The 3D view on cosmic baryon cycle

Aleksandra Hamanowicz

München 2020

The 3D view on cosmic baryon cycle

Aleksandra Hamanowicz

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Aleksandra Hamanowicz
aus Widminnen, Polen

München, den 27 Mai 2020

Erstgutachter: Prof. Dr. Simon White
Zweitgutachter: Dr. Klaus Dolag
Tag der mündlichen Prüfung: 13 Juli 2020

Kto sie zatrzyma – skamienieje, kto przystanie, zostanie przyszpilony jak owad, jego serce przebije drewniana igła, jego rece i stopy beda przedziurawione i przybite do progu i powały. (...) Ruszaj sie, ruszaj. Błogosławiony, który idzie. (...)

Nieważne, gdzie jestem, wszystko jedno, gdzie jestem. Jestem.

OLGA TOKARCZUK, *Bieguni*

This thesis has been carried out at the European Southern Observatory under the supervision of Dr Céline Péroux and Dr Martin Zwaan.

Contents

List of Figures

List of Tables

Zusammenfassung

Galaxien entwickeln sich nicht als isolierte Systeme - sie interagieren aktiv mit ihrer Umgebung. Dieser Gasaustausch ist ein wichtiger Aspekt in der Evolution von Galaxien und in Sternentstehungsprozessen. Frisches Gas wird von kosmischen Filamenten im intergalaktischen Medium über das Circum-Galactic-Medium (CGM) auf Galaxien übertragen. Galaktische Ausflüsse, ausgelöst von Supernovae, führen angereichertes Material zurück zum CGM. Ein Teil davon kehrt in Form von galaktischen Springbrunnen zurück zur Galaxie, ein anderer Teil wird vom galaktischen Wind aus dem System weggetragen. Um zu verstehen, wie Galaxien das zur Aufrechterhaltung der Sternentstehung erforderliche Gas auffüllen, untersuchen wir das durch CGM fließende Gas. Durch die Kombination der Emissions- und Absorptions-Tracer können wir die Eigenschaften der mit Quasar-Absorbern assoziierten Galaxien charakterisieren und ihre Wechselwirkungen mit Gas im CGM untersuchen. Wir verwenden eine detaillierte kinematische Analyse, um die Absorber mit den Galaxien im System zu verbinden.

Wir können diese riesigen gasförmigen Halos mit hellen, weit entfernten Punktquellen wie Quasaren untersuchen, indem wir die Metall- und Wasserstoffabsorptionslinien in ihren Spektren als Tracer des CGM der Vordergrundgalaxien identifizieren. Wir kennen Hunderttausende solcher Absorptionssysteme, aber nur eine Handvoll von ihnen haben identifizierte galaktische Gegenstücke, die wir untersuchen können. Um das Problem der Identifizierung und Charakterisierung von Absorberwirtsgalaxien anzugehen, haben wir das Multiwellenlängen-Projekt MUSE-ALMA-Halos erstellt. Wir finden, dass die meisten Absorptionssysteme Paaren oder sogar Gruppen von Galaxien zugeordnet werden können, die sowohl aus sternbildenden als auch aus nicht-sternbildenden Galaxien bestehen. Durch die Kombination der Emissions- und Absorptions-Tracer können wir die Eigenschaften der mit Quasar-Absorbern assoziierten Galaxien charakterisieren und ihre Wechselwirkungen mit Gas im CGM untersuchen.

Global gesehen tritt Gas, das durch das CGM strömt, in eine Galaxie ein und füllt die Gasreservoirs auf, die sich weiter in die molekulare Phase umwandeln - der direkte Brennstoff der Sternentstehung. Wir wissen, dass die Sternentstehung in dichten Molekülwolken stattfindet, jedoch ist die Verbindung zwischen dem allgemeinen molekularen Gasgehalt von Galaxien und der globalen Sternentstehung des Universums komplex. Die Geschichte der Sternentstehungsrate (Star Formation History - SFH) des Universums ändert sich mit der Zeit, und es ist nicht klar, was den $z \sim 2$ Höhepunkt und den anschließenden Rückgang antreibt. Neuere Studien haben einen möglichen Zusammenhang

zwischen der SFH und der kosmischen Dichte von H_2 - dem am häufigsten vorkommenden Molekül im Universum - angedeutet. Um das H_2 in Galaxien nachzuweisen, verlassen wir uns auf CO, das am zweithäufigsten vorkommende Molekül mit viel lichtstärkeren Übergängen, das mit ALMA bei Sub-Millimeter-Wellenlängen relativ einfach zu beobachten ist. Eine unverzerrte Möglichkeit, den Molekülgehalt von Galaxien über die kosmische Zeit hinweg zu untersuchen, sind ungezielte Vermessungen der Emissionslinien. Bisher waren solche Untersuchungen auf einzelne Felder beschränkt, die einen kleinen Teil des Himmels abdecken, was sie anfällig für die Auswirkungen der kosmischen Varianz macht.

Zur Ergänzung der Deep-Pencil-Beam-Durchmusterungen verwendeten wir einen alternativen Ansatz, indem wir im Rahmen des ALMACAL-Projekts archivierte Kalibrierungsdaten von ALMA verwendeten. Aus diesen Daten konnten wir eine ungerichtete Vermessung der CO-Emissionslinien konstruieren, die wir ALMACAL-CO nennen. ALMACAL zielt darauf ab, zahlreiche ALMA-Kalibrierscans für wissenschaftliche Zwecke zu verwenden. Dank der Einzigartigkeit des ALMACAL-Datensatzes können wir Galaxien über ein weites Gebiet untersuchen und sind nicht empfindlich gegenüber den Auswirkungen der kosmischen Varianz. Im ALMACAL-CO-Pilotprojekt haben wir die Felder mit den längsten Gesamtintegrationszeiten ausgewählt und elf Emissionslinien detektiert. Da für unsere Durchmusterung keine systematischen Zusatzdatensätze zur Verfügung stehen, haben wir keine unabhängigen Messungen der Rotverschiebung unserer entdeckten Galaxien und können daher nicht eindeutig bestimmen, welchen CO-Übergang wir gefunden haben. Um hier Abhilfe zu schaffen, nutzten wir die simulierte Spektrallinien-Energieverteilung der Galaxienpopulation aus den semi-analytischen Shark-Modellen, um die Nachweise mit dem entsprechenden CO-Übergang mit Hilfe eines probabilistischen Ansatzes zuzuordnen. Die Entwicklung der molekularen Gasmassendichte, die aus dem ALMACAL-CO-Pilotprojekt abgeleitet wurde, stimmt mit den Grenzen überein, die durch die früheren ungezielten Untersuchungen (ASPECS, COLDz) gesetzt wurden. Gleichzeitig stimmt sie auch mit dem Fehlen jeglicher Entwicklung der molekularen Gasmassendichte überein.

Von Untersuchungen der Gasströme und der Verteilung der Metalle in dem CGM bis hin zum molekularen Gasgehalt der Galaxien verfolge ich in meiner Arbeit die verschiedenen Phasen der Baryonen. Die lokale Ansicht des CGM von Galaxien erlaubt es, die Materieströme zwischen IGM und CGM sowie die Metallizität der Materie zu studieren, ein wichtiger Aspekt der Entwicklung von Galaxien. Schließlich folgt die Entwicklung der molekularen Gasdichte der Beziehung zwischen Baryonen und der Sternentstehungsgeschichte des Universums.

Abstract

Galaxies are not systems evolving in isolation, on the contrary, they are actively interacting with the surrounding medium. The local cycle of baryons through the flows of gas are an important aspect of the evolution of galaxies and star formation processes. Pristine gas is accreted from cosmic web filaments from the Intergalactic Medium (IGM) onto galaxies through the Circumgalactic Medium (CGM). Strong outflows from supernovae and AGNs expel enriched material to the CGM, some of which falls back in the form of galactic fountains, another part is carried away from the system by the galactic wind. To understand how galaxies replenish gas needed to sustain the star formation, we study the gas flowing through the CGM. By combining emission and absorption tracers we characterize the properties of the galaxies associated with quasar absorbers and study their relation with the gas in the CGM. We use bright distant point sources, such as quasars, in the spectra of which we identify hydrogen and metal absorption lines, to probe the CGM around foreground galaxies.

We know hundreds of thousands of such absorption systems, but only a few dozen of them have identified and well-studied galactic counterparts. To address the problem of identification and characterisation of the absorber host galaxies I built the multi-wavelength survey known as MUSE-ALMA Halos. The survey consist of the five quasar fields observed by MUSE, with known 14 $z < 1$ absorbers. I found that most absorption systems can be associated to pairs or even groups of galaxies, which members consist of both star-forming and quiescent galaxies. Combining emission and absorption I characterized the properties of the galaxies associated with quasar absorbers and study the gas in their CGM.

The cycle of baryons influence the galaxy evolution also globally. Gas travelling through the CGM enters a galaxy and replenishes the gas reservoirs which further transform into the molecular phase - the direct fuel of star formation. We know that star formation takes place in dense molecular clouds, however, the link between the general content of molecular gas of galaxies with the global star formation of the Universe is complex. The star formation rate history (SFH) of the Universe changes with time and it is not clear what drives the $z \sim 2$ peak and subsequent decline. Recent studies have suggested a possible link between the SFH and cosmic density of H_2 - the most abundant molecule in the Universe. To trace the H_2 in galaxies we rely on CO, the second most abundant molecule with much brighter transitions, relatively easy to observe at sub-millimetre wavelengths with ALMA. An unbiased way to study the molecular content of galaxies across cosmic time is through untargeted emission-line surveys. To date, such surveys have been limited to single fields,

covering a small part of the sky which makes them susceptible to the effects of cosmic variance.

To complement the deep pencil beam surveys I refered to an alternative approach, using ALMA archival calibration data from the ALMACAL project. From these data, we were able to construct an untargeted CO emission line survey which I called ALMACAL-CO. ALMACAL aims at using numerous ALMA calibration scans for scientific purposes. Thanks to the uniqueness of the ALMACAL dataset we can study galaxies over a wide area and are not sensitive to the effects of cosmic variance. With the use of the calibrator fields, ALMA can, in effect, be transformed into the survey instrument without additional costly time allocation for the observatory. In the ALMACAL-CO pilot, we chose the fields with the longest total integration times and detected eleven emission lines. Since for our survey no systematic ancillary datasets are available, we have no independent measurements of the redshift of our detected galaxies and hence cannot make a unique identification of which CO transition we have found. To remedy this, we made use of the simulated spectral line energy distribution from the population of galaxies from the Shark Semi-Analytical Models, in order to associate the detections with the corresponding CO transition using a probabilistic approach. The evolution of the molecular gas mass density derived from the ALMACAL-CO pilot is in agreement with the constraints put by previous untargeted surveys (ASPECS, COLDz). However, taking into account the large statistics unceratinties, my current measurement is also still consistent with no with the lack of any evolution of molecular gas mass density. To provide further constraints on the molecular gas mass density evolution, we will extend the ALMACAL-CO pilot survey to whole ALMACAL database, expanding the probed volume 50 times. With this extension we expect to detect over 100 CO emitters.

From studies of the gas flows and the distribution of metals in the CGM to the molecular gas content of the galaxies, in my work, I traced the different phases of baryons. The local view of the CGM of galaxies allows for studies of the flows of matter between IGM and CGM as well as the metallicity of the matter, an important aspect of the evolution of galaxies. Finally, the evolution of the molecular gas density follows the global relation between baryons and the star formation history of the Universe.

Chapter 1

Introduction

Baryons constitute of only about 5 per cent of the total matter budget in the Universe, nevertheless, studies of their interaction and structure formation dominate modern astrophysics. The evolution of galaxies together with the star formation processes are complex problems addressed by observation of objects at different evolutionary stages, and importantly also by high-resolution observations addressed by observing objects on small, local scales. In parallel, hydrodynamical simulations continuously improve reproducing the observational properties of galaxies across redshifts and our understanding of the physics of the baryonic matter.

The main topics in today's extragalactic astrophysics are the evolution of star formation of galaxies, the processes driving quenching and the role that the gas supply plays in these processes. We look at the cycle of baryons on the two scales. Globally the evolution of the cosmic density of baryons and how their supply to galaxies influences the star formation history of the Universe. Locally, the flows of matter regulate the star formation on the scales of individual galaxies, replenishing the Interstellar Medium with gas and polluting the metal-rich material expelled by supernovae or AGN activity. In the following sections, I introduce these problems, where the cosmic evolution of baryons plays a substantial role, and describe the methods that I have employed to address these questions.

1.1 The local baryon cycle infuence on the evolution of galaxies

The evolution of galaxies and their transformation from star-forming to quenched systems is regulated by the flows of the gas supply. The local cycle of baryons between the galaxy Interstellar Medium (ISM) and Circumgalactic Medium (CGM) regulates the star formation processes, provides the supply of new material and distributes the newly formed metals. To sustain the star formation galaxies cannot evolve in isolation, but must replenish the reservoirs of gas in their ISM by accretion from the Intergalactic Medium (IGM). The star formation and AGN activity, in turn, launch galactic winds, expelling metal-rich material back to the CGM and possibly also removing it completely from the system, af-

fecting the available gas reservoirs in the ISM and potentially preventing further accretion, thus initiating quenching. That cycle of gas plays a crucial role in the processes driving galaxy evolution and star formation. Studies of the Circumgalactic Medium described in this section can answer the questions of how galaxies sustain star formation, how much matter is accreted through CGM how much matter is expelled in form of the galactic winds and what is the role of these outflows in the quenching transformation.

1.1.1 The Circumgalactic Medium

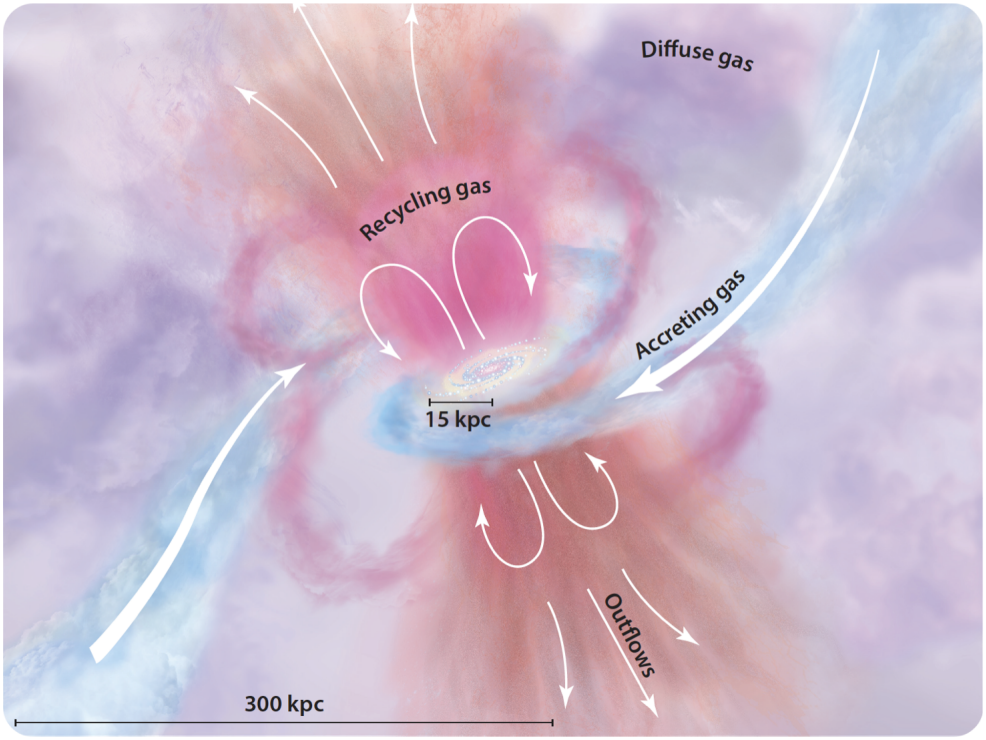

Figure 1.1: The cartoon view of the Circumgalactic Medium (illustration from Tumlinson et al., 2017). The blue represents metal-poor gas inflowing from the IGM and accreted through the CGM onto the galaxy. The star-formation enriched material (orange, pink) is expelled from the galaxy in the form of galactic winds, part of which falls back onto into the ISM as galactic fountains. The purple colour of the surrounding medium represents the mixed diffuse gas filling the CGM.

The Circumgalactic Medium (CGM) is a loosely defined gaseous halo surrounding a galaxy, spanning up to about its virial radius($R_{\rm vir}$ ∼300 kpc for the L_* galaxy). The CGM is the arena of gas inflows from the Intergalactic Medium and recycling of metals expelled from the galactic disc through the galactic winds launched by supernovae of strong star formation activity (Figure 1.1). Most of the material remains bound and falls back in the form of a galactic fountain (Fraternali, 2017), enriching the ISM with metals. It is not clear

which fraction of material is expelled from the galaxy in galactic winds, which fraction of metals are recycled back to the ISM and what are the typical timescales of these processes.

The CGM is dominated by cool clumpy collisionally ionised ($T : 10^{4-5.5}$ K) gas of low density, with velocities lower than the galaxy escape velocity. It is a multi-phase medium, a mix of hot ionised gas (traced by O VI) with layers or clouds of cool gas within the $R_{\rm vir}$. The cold gas density profile drops with distance from the galaxy, while the hot gas remains mostly constant gradually transforming to diffuse and hot ($T : 10^6$ K) Warm-Hot Intergalactic Medium (WHIM) expanding to the 1 Mpc from the galaxy and observable in X-rays (Tumlinson et al., 2017).

One of the challenges of studying the CGM is its low surface brightness which makes it challenging to detect with present-day instruments (Wisotzki et al., 2018; Augustin et al., 2019). Additionally, most of the metal and hydrogen emission lines lie in UV, making the observations challenging, often requiring the use of space UV observatories.

To study the CGM we use the absorption signatures in the spectra of distant bright sources or, alternatively, the so-called down-the-barrel technique when we use the light of the host galaxy to trace the absorption lines from its own CGM. The CGM traced by the quasar absorbers has been discovered around different types of galaxies: quasars, star-forming galaxies at different redshifts and also around quiescent objects. The presence of the cool medium around quiescent objects raises the question of the nature of quenching processes blocking that material from being accreted onto the galaxy (Tumlinson et al., 2017).

1.1.2 Quasar absorbers

The main method for studying the CGM today is through intervening quasar absorber system found in the UV and optical spectra of background quasars. The quasar absorbers are a unique method of tracing intergalactic medium gas as the detection of the absorption depends only on the brightness of the background source, making it an effective method to trace the evolution of the CGM through time (Figure 1.2). Absorption studies are sensitive to the low column densities system and probe a range of gas properties such as temperature, density or ionisation state. This method, however, grants only a single pencil beam sight-line through a gaseous halo, not enough for its complete characterisation. Thanks to the COS Hubble (Tumlinson et al., 2013; Peeples et al., 2014) and ground-based surveys like KBSS (Steidel et al., 2014) we know tens of thousands of quasar absorbers and deduce statistically the general properties of the CGM around star-forming galaxies.

We distinguish several types of absorbers, based on the column density of the HI absorption (N(HI)) in units of atoms cm^{-2}. The main method to obtain the column density from the absorption line is through fitting the Voigt profile, representing the physical line broadening processes (a convolution of thermal effect represented by the Gaussian profile and collisional broadening described by the Lorentzian profile) and described by the doppler parameter b. The function of the column density of the probed by the absorption is proportional to the equivalent width of the line and depends on the column density regime which it is probing. Their relationship is described by the curve of growth.

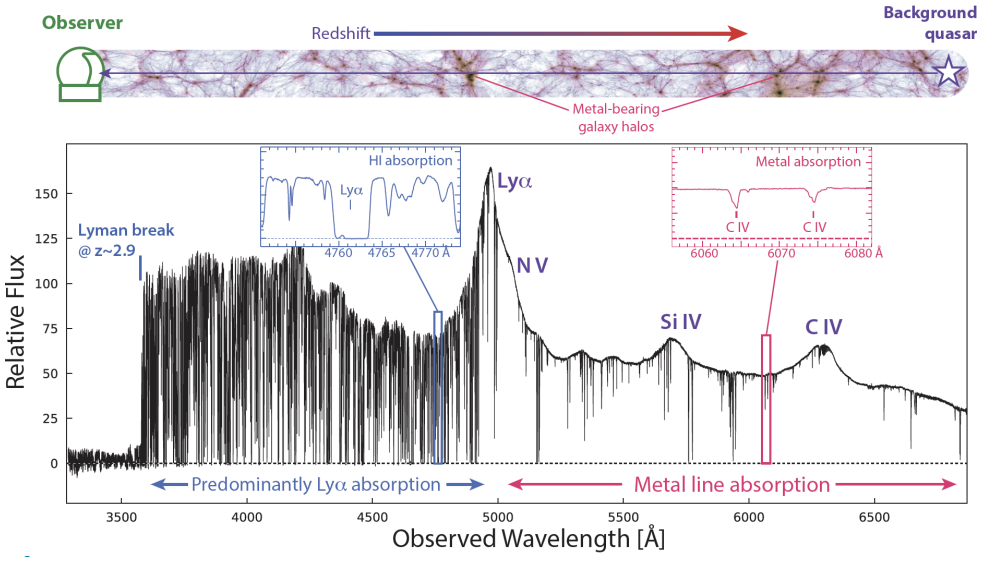

Figure 1.2: An illustration of the quasar absorberption technique (illustration from Péroux & Howk, 2020). The signature of the foreground material gets imprinted as absorption lines in the spectrum of the distant background quasar. The Lyα forest is the signature of the gas in IGM, while the broad, often saturated Lyα absorption together with metal lines originate mostly in the CGM of galaxies.

Absorption of $\log(N(\mathrm{HI})) \sim 15$ cm^{-2} is a limiting case between IGM Lyα forest and CGM and the features can be fitted with Voigt profiles. In the low-column density regime, the column density is linearly proportional to the equivalent width of the line. At the column density of $\log(N(\mathrm{HI})) \sim 16$ cm^{-2}, some absorption lines become saturated and equivalent width and b parameter become degenerate - such systems are called partial or complete Lyman Limit Systems (LLS). At $\log(N(\mathrm{HI})) \sim 18$ cm^{-2} the Lyman series is completely saturated and the fitting is possible only for metal lines. With a growing probed column density the relation between the equivalent width and column density becomes also more complicated firstly becoming logarithmic and the as a square root of the column density. The nature of the highest column density absorbers, sub-DLA ($\log(N(\mathrm{HI})) > 19.0$ cm^{-2}) and DLAs (Damped Lyman Alpha systems, $\log(N(\mathrm{HI})) > 20.3$ cm^{-2}) is still not fully understood, but most likely connected to galaxy discs.

In the CGM gas, which is in constant movement, we can trace through absorption the accretion of new material onto the galaxy disc as well as the outflowing gas in the form of galaxy wind. However, a single sightline view through that gaseous halo makes the disentangling of these gas flow challenging. Search for the inflows and outflows of gas are then done statistically, by looking at the metallicity gradient between the absorbing gas and galaxy, or tracing the metal distribution as a function of the azimuthal angle - the angle between the galaxy disc and the quasar sightline (Bordoloi et al., 2011; Bouché et al., 2012; Kacprzak et al., 2012). We expect inflows to be metal-poor (lower metallicity than

the galaxy's ISM) while outflow to be metal-rich. The distribution of the LLS metallicity (Lehner et al., 2013; Wotta et al., 2019) was found to show such double-peak features - low and high metallicity attributed to the tracers of inflow and outflow. However, this bimodality is not reproduced by the hydrodynamical simulations (FIRE/EAGLE, Hafen et al., 2017). Later, gas mixes in the internal regions of the CGM, blurring the metallicity and azimuthal angle dependence (Prochaska et al., 2017) leading to an unimodal distribution of metals within $r \sim 160$ kpc from a galaxy.

Obtaining metallicity measurement of the absorbing gas is difficult and requires simultaneous measurements of metals and Lyα absorption lines. Additionally, from the studies of the reddening of the background quasar, we know that the CGM also contains dust, as reddening observed in the ISM, extends also to the hundreds of kpc away from the galaxies (York et al., 2006; Ménard et al., 2010). In metallicity estimates, one must then account for the different degrees of dust depletion of different metal species (De Cia, 2018).

Altogether CGM studies provide a powerful method of measuring the evolution of neutral gas across redshifts as well as understanding the gas flows, processes playing a vital role in the evolution of galaxies.

1.1.3 The galaxy hosts of quasar absorbers

With the tens of thousands of absorption systems known only a few dozen of them have galaxy counterparts. Historically, the search for galaxy counterparts was limited to the photometric surveys, where typically the galaxy closest in projection to the sightline was assumed to be the host. Subsequently it's redshfit was confirmed spectroscopically (Bergeron et al., 1988; Boisse et al., 1998). Thanks to a new generation of spectrographs, the integral field unit spectrograph (IFU), especially MUSE, searching for the galaxy counterparts has become more efficient. IFU is a combination of an image and a spectrograph, with each pixel harbouring a spectrum. MUSE, with its 1 arcminute field of view allows for a search for galaxies up to 150–300 kpc at $z = 0.3 - 1$ from the QSO sightline. The DLA hosts are found to follow the mass-metallicity (Christensen et al., 2014) and metallicity-luminosity relations (Krogager et al., 2015), and are more likely to be detected if they are massive metal-rich galaxies, which form only a subset of the DLA hosts. With to the large field of view and ability to spectroscopically classify all continuum and emission line sources, it was possible for the first time to look for multiple galaxies associated with quasar absorbers (e.g. Bielby et al., 2017; Klitsch et al., 2018; Péroux et al., 2019, Chapter 2).

A new view of the CGM hosts also come from the discovery of the bright CO emission from some of the DLA counterparts (e.g. Neeleman et al., 2016). The CO emission lines offer a powerful tool to detect the host galaxies of quasar absorbers at high redshifts ($z > 2$) but are usually selected by their high metallicity (Neeleman et al., 2018). CO can be also detected in the members of the galaxy groups at low redshift ($z < 1$, Klitsch et al., 2018; Péroux et al., 2019), exposing the large reservoirs of molecular gas but long depletion timescales. These findings highlight that galaxies selected based on the presence of the quasar absorption are inefficient in the converting their gas into stars.

Combining a multi-wavelength view of the quasar absorbers and their host galaxies allows to trace the evolution of metals and HI gas and to understand the gas flows in and around galaxies, the important processes of gas replenishment in the galaxy ISM and the potential sources of galaxy quenching.

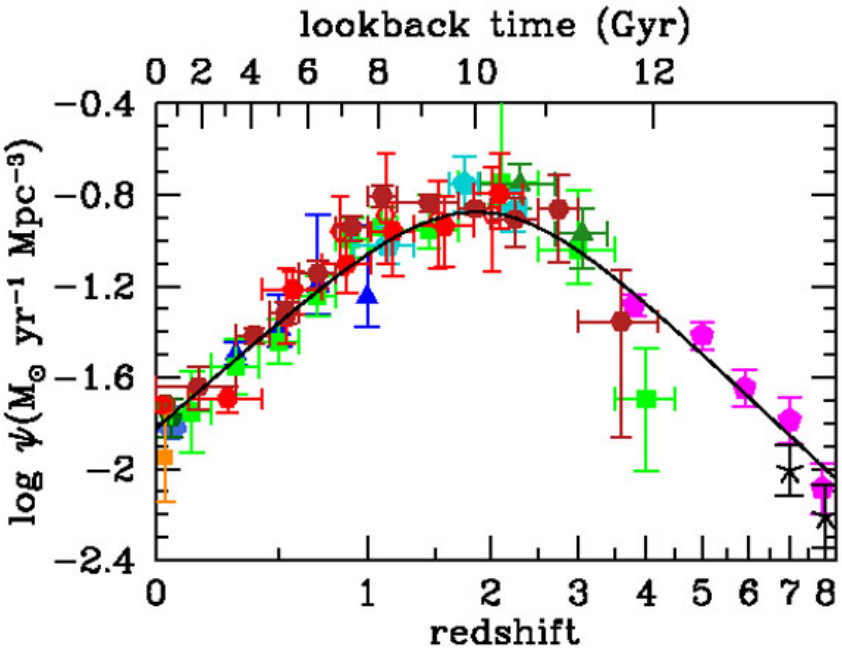

Figure 1.3: The star formation history of the Universe diagram from Madau & Dickinson (2014). The measurement points are the compilation of the FUV and IR rest-frame observations, with each singular star formation value derived from the integration of the corresponding luminosity function. The solid line represents the best fit for the star formation history.

1.2 The global evolution of baryons and relation with the star formation history

The evolution of the star formation history of the Universe is well established observationally up to the epoch of reionisation by the measurements of the star formation rates in objects across redshifts. Madau & Dickinson (2014) report the compilation of multiple

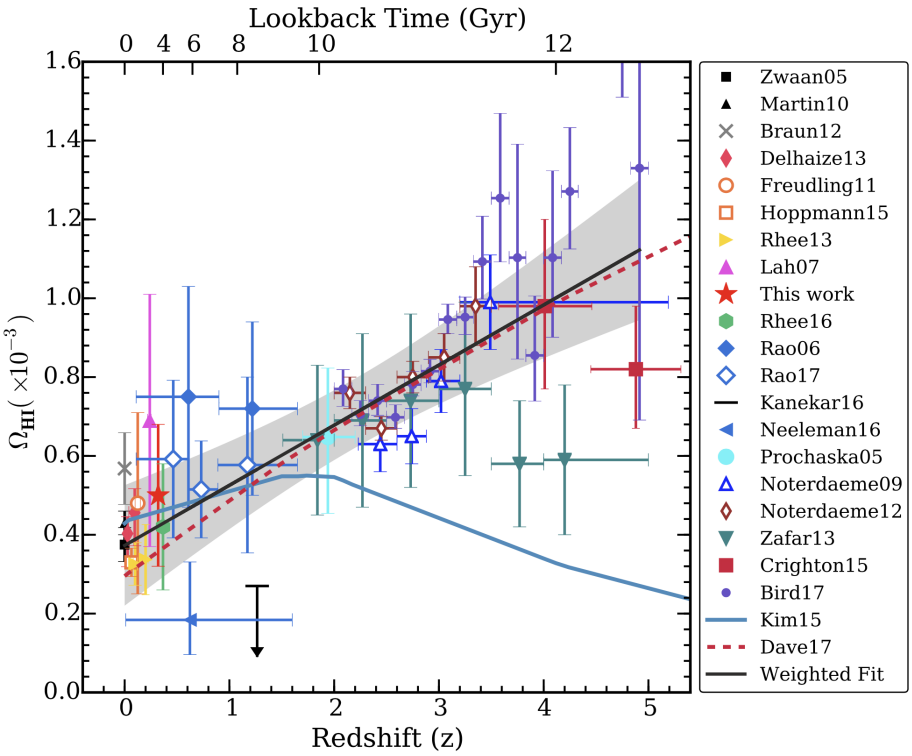

Figure 1.4: The cosmic neutral gas density as a function of redshift and look-back time (figure from Rhee et al., 2018). The multiple points are a compilation of the measurements of HI density from the literature (references indicated in the legend), consisting of different techniques including 21-cm emission lines surveys, 21-cm emission line stacking of optically identified galaxies and Lyα measurements in quasar absorbers. The blue solid line shows predictions from the semi-analytical models, red dashed line shows the predictions from MUFASA simulations (references in the legend). The black solid line with the shaded area is the weighted fit to the measurements and the 95 per cent confidence interval. The neutral gas density evolves as a function of redshift as $(1+z)^{0.57}$.

UV and IR measurements (Figure 1.3), leading to a complete view of the star formation history of the Universe (SFH). The cosmic star formation activity peaked about 3.5 Gyr after the Big Bang and declined by a factor of 10-15 until the present day. The ensemble of observations of galaxies in the high-z Universe are sparse and the exact slope of the rise of star formation since the epoch of reionisation to the $z \sim 2$ peak is still debated. We do not understand what drives the evolution of the global star formation. One of the possible explanations implies a close tie to the evolution of the cosmic cold gas reservoir (mainly HI and H_2).

At redshift $z = 0$, the star formation is an inefficient process, with the efficiency of converting gas mass to stars of only a few per cent (Zuckerman & Evans, 1974; Krumholz & Thompson, 2007). To sustain their star formation, the galaxy's ISM gas reservoirs must

be constantly replenished with the gas filling the Intergalactic Medium. The cosmimc evolution of neutral hydrogen is thought to potentially trace the changes in the star formation history of the Universe.

To initiate the process of star formation, gas in the ISM must cool and transform into the molecular phase. The molecular gas-phase proceeds the star formation (Leroy et al., 2008; Schruba et al., 2011) as any other ISM phases cannot form stars directly unless they cool sufficiently to form molecules. Measuring the molecular gas mass is crucial to understand the star formation in a galaxy and potentially can explain the evolution of the star formation history of the Universe.

The two phases of cold gas (atomic and molecular) are both important actors in sustaining star formation and play a role in the evolution of the global star formation across cosmic time. In the next sections, I discuss the observations of both gas phases and their relation to the star formation history.

1.2.1 The evolution of the atomic gas

The neutral gas, dominated by the atomic hydrogen (HI), fills the Inter-Galactic medium and is gradually accreted by the star-forming galaxies (Fox & Davé, 2017). Hydrogen is traced through several observational techniques. In the local Universe, the radio observations of 21-cm HI emission and absorption map HI locked in the ISM of galaxies and their immediate surroundings. However, due to the low brightness of the 21-cm emission line, its detection in galaxies is only possible up to redshift $z < 0.2$ (Rhee et al., 2018, and references therein), and is extended to redshift $z \sim 0.4$ through stacking the spectra of multiple galaxies (Kanekar et al., 2018).

To measure the neutral gas mass at higher redshifts one makes use of observations of the Lyα line in intervening quasar absorbers (such as DLA, see Section 1.1.2). The detectability of quasar absorbers depends only on the brightness of the background source, allowing for measurements of Lyα absorption up to high redshifts $z \sim 5$ (e.g. Neeleman et al., 2016).

Thanks to compilation of HI observations, the evolution of the cosmic density of neutral gas is well determined up to $z \sim 5-6$ (Rhee et al., 2018, Figure 1.4). The comoving mass density of neutral gas evolves little over time, decreasing with redshift as $(1+z)^{0.57}$ since redshift $z \sim 5$ (Péroux & Howk, 2020, ARAA, in press). Although the HI gas is the main constituent of the ISM of star-forming galaxies, its global evolution does not mirror the evolution of the star formation history and cannot explain its dramatic change between $z = 2$ and $z = 0$.

1.2.2 The molecular gas in galaxies across redshifts

From the studies of the Milky Way and other nearby galaxies we know that stars form in the dense cold dusty molecular clouds (Giant Molecular Clouds, McKee & Ostriker, 2007; Kennicutt & Evans, 2012). In these conditions, gas can reach the Jeans criterion and collapse into a protostar. In the high gas densities of the clouds, molecules are protected

from dissociation by the UV radiation by self-shielding and are additionally shielded by dust (Tielens, 2005).

The most abundant molecule in the Universe is the molecular hydrogen, H_2. The molecule is symmetric, hence the only available transitions are the faint vibrational ones. Their excitation requires a temperature of $T_{\rm exc} = 500$ K, while the temperature of a typical molecular cloud is much lower, $T = 10 - 100$ K. The H_2 molecule cannot be excited in a typical molecular cloud.

The typical tracer of the molecular gas is the second most abundant molecule, carbon monoxide - CO (Carilli & Walter, 2013). CO is located deeper into the clouds as on the surface it is dissociated by the UV radiation. Its excitation temperature ($T_{\rm exc} = 5$ K) lies well within the conditions of the Giant Molecular Clouds, making CO a good tracer of the star-forming regions. Additionally, its low-J CO transitions have high luminosities and can be observed by sub-mm interferometers up to high redshifts.

CO emission is driven by the gas heating from the dust radiation. A few per cent of the heating goes to the excitation of another bright tracer of star formation, the [CII] 158 μm line. The [CII] emission line is even ten times brighter than low-J CO transitions and its emission from high-z galaxies ($z > 4$) is observed at the sub-millimetre frequencies.

1.2.3 Cosmic molecular gas estimates

To transform the observations of the CO luminosity to the molecular gas mass we have to scale the integrated line luminosity with the CO-to-H_2 conversion factor $\alpha_{\rm CO}$ (Bolatto et al., 2013). The $\alpha_{\rm CO}$ depends on several parameters: the cloud density, the temperature of the transition and especially the ISM metallicity (Leroy et al., 2011; Narayanan et al., 2012; Feldmann et al., 2012; Genzel et al., 2012). The most common values vary between $\alpha_{\rm CO} = 3.6$–4 for a Milky Way like galaxy and $\alpha_{\rm CO} = 0.8$ for starburst galaxies. Most measurements (including the projects presented in this thesis) assume the Milky Way value of the conversion factor. The uncertainty of $\alpha_{\rm CO}$ and its evolution with redshift and ISM conditions is one of the main challenges in the extragalactic sub-millimetre astrophysics today.

The second challenge is the uncertainty of the Spectral Line Energy Distribution (SLED) - the relation between the fluxes of the different molecular transitions (Figure 1.5). The estimate of the molecular gas mass depends on the measured luminosity of the CO(1−0) line. However this transition is only observable with ALMA up to the redshift $z \sim 0.5$, or above redshift $z > 2$ with the VLA but only for the brightest sources. In most of the cases, the only observable CO lines are higher-level transitions and the conversions between their luminosities and the CO(1−0) luminosity requires a knowledge of the SLED of that galaxy.

The relative luminosities of the CO excitation levels depend strongly on the ISM properties. The SLEDs have been mapped in detail only for several sources including the Milky Way, AGNs, quasars and several sub-millimetre dusty galaxies (SMGs) (Carilli & Walter, 2013). Measurements of multiple CO transitions in a typical star-forming galaxy require a long observing time and such systems are poorly sampled. The most representative pre-ALMA CO SLEDs of star-forming galaxies were measured by Daddi et al. (2010, 2015)

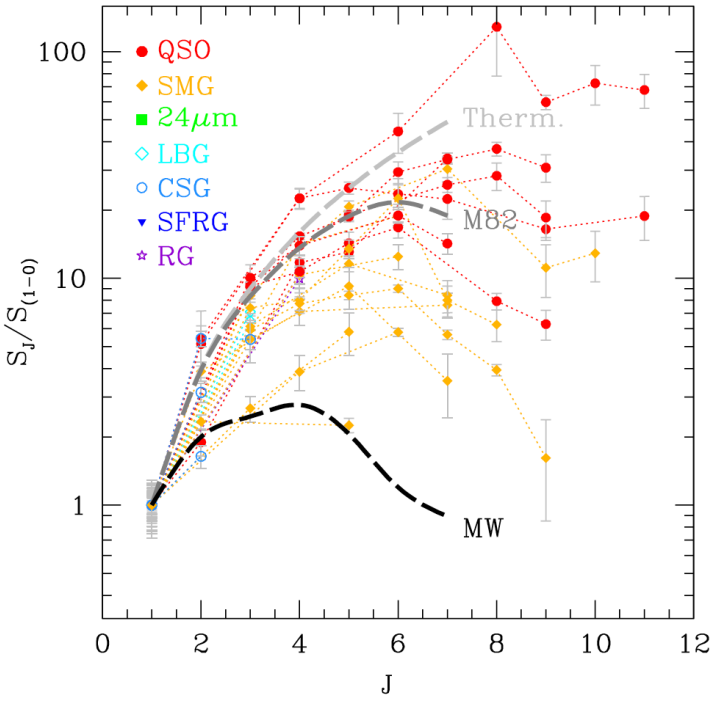

Figure 1.5: The CO Spectral Energy Distribution measurements for several types of galaxies, including quasars, sub-millimetre dusty galaxies and limited measurements from star-forming galaxies (illustration from Carilli & Walter, 2013). Dashed lines mark the SLEDs of Milky Way, nearby starburst M82 and constant brightness temperature on the Rayleigh-Jeans scale. The measurements of the star-forming galaxies are limited and their SLEDs are not well sampled. The best-sampled quasars and SMGs show much higher excitation of the CO emission lines then one measured in Milky Way. Understanding of the SLEDs in different galaxies and the evolution with redshift is important for precise measurements of molecular gas content of galaxies.

in four $z \sim 2$ BzK galaxies (galaxies selected through their 4000 Å breaks). Quasars and SMGs, with the most excited ISM, have steep SLEDs, much different from observed in the Milky Way. BzK galaxies have also more excited ISM than that of the Milky Way, however, their exact shape above $J_{\rm up} = 4$ is unknown. With these uncertainties, the most commonly used conversion between high-J lines and CO(1−0) is the Milky Way CO SLED coefficients. Together with the unknown relation between the CO and H_2 (Walch et al., 2015), poorly sampled CO SLEDs introduce another uncertainty in the molecular gas mass measurements.

1.2.4 Targeted and untargeted sub-millimetre emission line surveys

Measuring the molecular gas in star-forming galaxies is made through targeted or untargeted surveys. In targeted surveys, a sample of galaxies with well-established parameters (e.g. stellar mass, star formation rate) is chosen based on the luminosity or mass selection from the optical surveys like SDSS, CANDELS, GOODS etc. The chosen objects are then targeted for their CO emission lines (Tacconi et al., 2010; Daddi et al., 2010; Tacconi et al., 2018). Such observations allow for the characterisation of the certain sample of galaxies, preferably massive actively star-forming objects and measurements of their molecular masses, gas depletion times etc. Such studies established scaling relations between the molecular mass content of galaxies, their masses and SFR, the depletion time with redshifts etc. In such surveys, however, only a subset of the most massive galaxies is chosen with the assumption that such main-sequence star-forming galaxies are dominant reservoirs of molecular gas in the Universe. They demonstrate that their properties are alike the general population.

The targeted low redshift objects are dominating galaxy assemblies and harbour massive reservoirs of molecular gas, while their star formation rates are lower than observed in SMGs or starburst galaxies. These results suggest that the efficiency of star formation is low even in the presence of rich reservoirs of molecular gas. At $z > 1$ (PHIBSS, Tacconi et al., 2018), targeted star-forming galaxies exhibit gas-dominated, rotating discs, similar to spirals at $z = 0$ but with much higher SFR than reported in their analogues today.

The complementary approach is to conduct emission-line surveys, deep field scans analogue to the optical studies for emission lines in one field, preferably with available additional multi-wavelength data. In such an approach one can find known objects as well as ones previously not classified, without any bias on galaxy properties. So far, several such studies measured the cosmic density of molecular mass across redshift: the ALMA-based ASPECS in the Hubble Ultra Deep Field (Decarli et al., 2016, 2019) mostly from $z = 0$ to $z = 3$ and the VLA-based COLDz in the GOODs North and South fields (Riechers et al., 2019), reaching a redshift of $z = 3 - 6$ and the search for secondary detections in the fields of PHIBSS-2 survey (Lenkić et al., 2020). The measurements report evolution of molecular gas density mirroring the star formation history of the Universe with a clear maximum at $z \sim 2$ and subsequent decline towards $z = 0$. However, the untargeted measurements are still based on a small number of detections and the molecular mass estimations are plagued by large uncertainties. Additionally, the local measurements of the ASPECS survey do not agree with the xCOLD GASS results of redshift $z = 0$ star-forming galaxies (Saintonge et al., 2017) pehaps due to the strong effects of cosmic variance on the measurements.

1.2.5 The evolution of the molecular gas

The molecular gas, especially the most abundant molecule H_2, is directly connected to the star formation in contrast to the atomic hydrogen. The molecular gas content of a galaxy can be measured through the multiple CO emission lines observable in FIR and in the

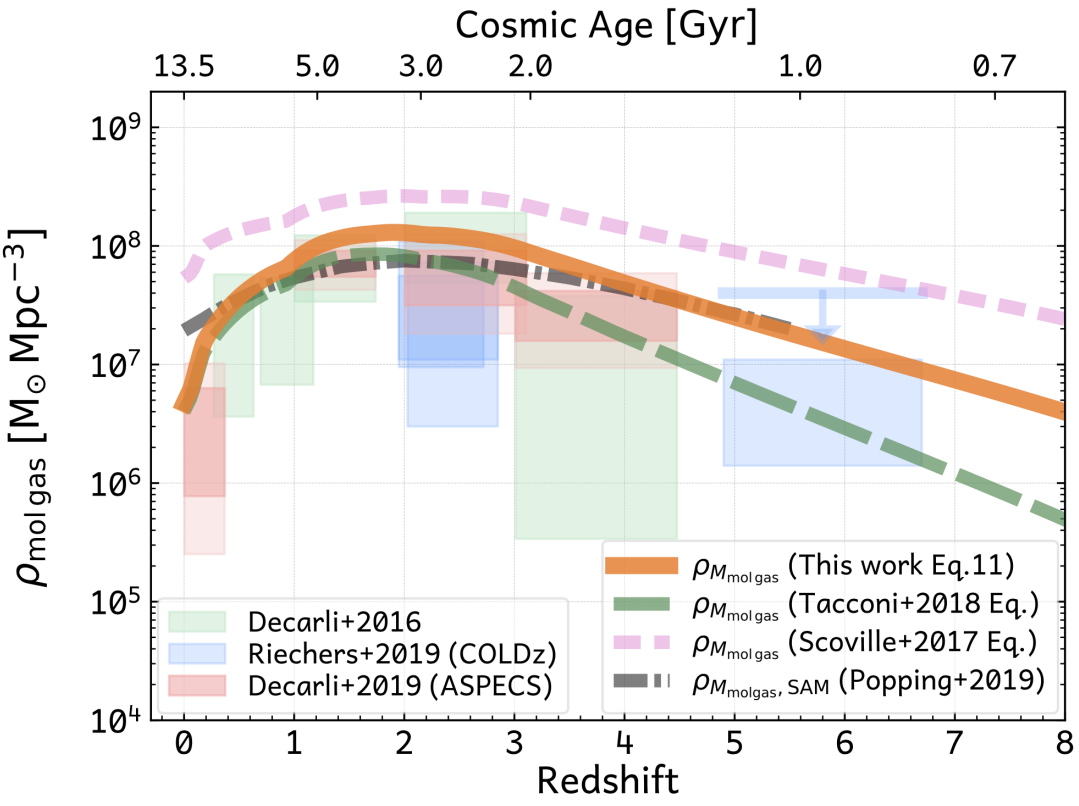

Figure 1.6: A compilation of the measurements of the evolution of molecular gas density with redshift (illustration from Liu et al., 2019). The boxes represent results from the untargeted emission-line surveys (ASPECS – Decarli et al. (2016, 2019), and COLDz – Riechers et al. (2019). The solid lines represent the scaling relations from Tacconi et al. (2018) (green), Liu et al. (2019)(orange) and dust-based estimates from Scoville et al. (2017). The grey line shows the prediction from the Semi Analytical Models in Popping et al. (2019).

sub-millimetre frequencies.

Limited simultaneous sub-millimetre frequency coverage and small fields-of-view of sub-millimetre interferometers make surveying the molecular emission challenging and the constraints on the evolution of the molecular gas are not yet comparable in accuracy to the precision of HI measurements. Molecular gas observations in galaxies at different redshifts are conducted through two approaches: by targeting massive star-forming galaxies or by deep scans of parts of the sky in search of molecular emission lines. Scaling of the observed line luminosities to the molecular gas mass is also difficult due to the uncertainties in the CO-to-H_2 conversion factor α_{CO} and the unknown Spectral Line Energy Distribution (SLED) of the CO emission ladder. The results of the CO emission lines so far, report that the shape of the evolution of molecular gas mass resembles the SFH with a significant peak of the molecular gas mass density at redshift $z = 2$ and a drop of a factor of two to

redshift $z = 0$ (Figure 1.6).

Globally neutral gas dominates molecular gas at all redshifts by at least a factor of 2–3. The molecular gas mass density, however, does not evolve as dramatically as the SFH. Additionally, the untargeted observations at redshift $z < 1$ suffer from the fact that only small cosmological volume are probed and might not be representative of the Universe and require further observational studies.

1.3 Summary

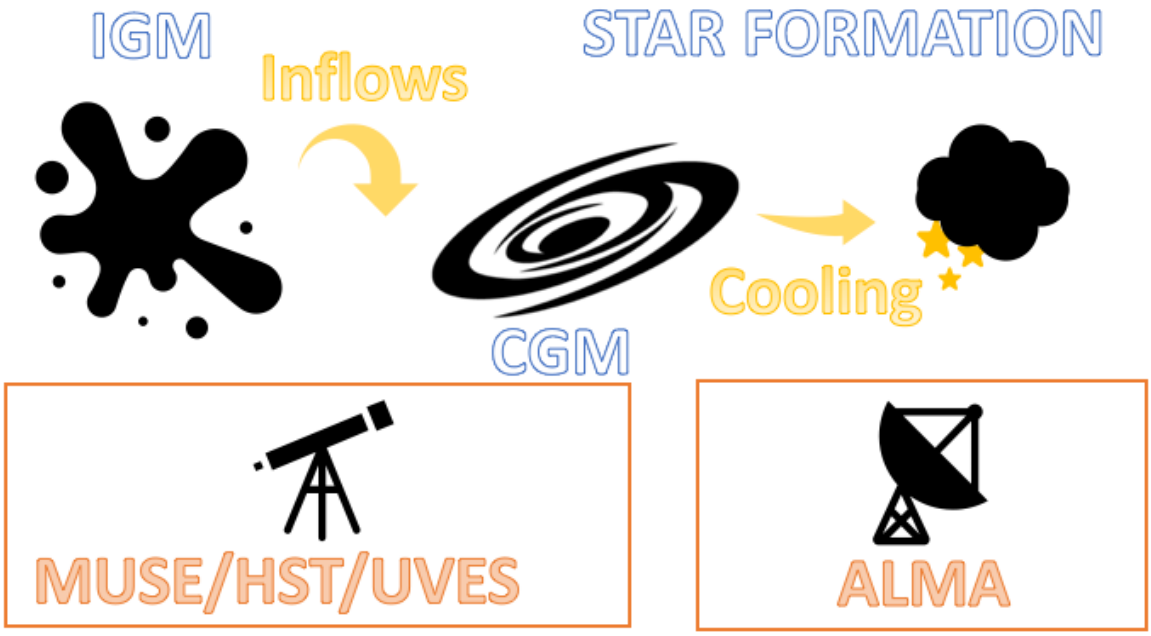

Figure 1.7: The cartoon illustration of the connection between the projects presented in this thesis. From the diffuse gas in the IGM, through the inflows by the CGM gas replenishes the galaxy ISM allowing for the star formation. Cooling down it forms molecules, and in the heart of the Giant Molecular Clouds forms new stars. With multiwavelength observational techniques, we can address the next steps of that gas flows process and address related questions of baryon cycle both locally and globally.

The question of the cause of the evolution of the star formation history of the Universe and the impact of baryon flow on the evolution of galaxies are the main motivations of the projects described in this thesis (summary, Figure 1.7). We study the gas flows through the CGM of $z < 1$ galaxies with the MUSE-ALMA Halos survey described in Chapter 2. To complement the untargeted molecular gas surveys we created the ALMACAL-CO survey (Chapters 3 and 4), utilising ALMA calibration data to search for molecular emission lines. The approach of using multiple fields for the survey potentially overcomes the issues of the

cosmic variance, problematic for the untargeted surveys conducted so far. In the end, I summarise both projects in Chapter 5 and present prospects for this work and extragalactic astrophysics in general.

Chapter 2

MUSE-ALMA Halos: physical properties and environment of $z \leq 1.4$ HI quasar absorbers

The content of this chapter was presented in the published article Hamanowicz et al. (2020), MNRAS, 492, 2347.

2.1 Introduction

Galaxies grow, evolve and sustain their star formation thanks to the accretion of gaseous material from filaments of the cosmic web (Martin et al., 2012; Rubin et al., 2012). In turn, strong outflows from supernovae or AGN feedback expel enriched material out of the galaxies (Shull et al., 2014), some of which comes back in the form of galactic fountains (Fraternali, 2017), while some is carried away from the system by galactic winds. Most of the gas recycling and inflow happens within a few hundreds of kiloparsecs from the galaxy centre in a region dubbed the Circumgalactic Medium (CGM). These extended gaseous halos surrounding galaxies are transition zones where inflowing Intergalactic Medium (IGM) material meets metal-enriched gas expelled from galaxies (Tumlinson et al., 2017).

Investigating the physical processes taking place within the CGM is key to understanding galactic evolution. However, studies of these extended regions in emission are challenging due to the low surface brightness of the gas (Frank et al., 2012; Corlies et al., 2018; Augustin et al., 2019). So far emission from the CGM has only been detected around extreme ionization sources, like luminous quasars, in the form of gigantic filamentary Lyα nebulae (Cantalupo et al., 2014; Borisova et al., 2016; Arrigoni Battaia et al., 2019; Lusso et al., 2019; Umehata et al., 2019) and using deep MUSE exposures, around faint galaxies at redshift $z = 3 - 6$ (Wisotzki et al., 2018). To study the CGM around typical star-forming galaxies, we must rely on distant bright point-like light sources such as quasars or Gamma Ray Bursts whose light gets attenuated by the material around a foreground galaxy (Bouché et al., 2007; Péroux et al., 2011; Christensen et al., 2017). Various ab-

sorption lines, from the most abundant Lyα, to numerous metal lines like MgII, FeII or CIV, originating in the intervening systems can reveal information about kinematics and metallicities of the absorbing medium. Metals can, in principle, be used as tracers of gas flows in the CGM. Lehner et al. (2013), Quiret et al. (2016) and more recently Wotta et al. (2019), observed a bimodality in the metallicity distribution of the low redshfit quasar absorbers, possibly a trace of metal-poor inflows and metal-rich outflows.

Understanding the physical properties of the CGM is at present best constrained observationally by the identification of the host galaxies linked to the gas observed in absorption. Historically, the first associated galaxy candidates have been identified on wide-field images of the quasar field as the closest object to the quasar sight-line (Bergeron et al., 1988; Bergeron & Boissé, 1991; Steidel et al., 1994; Le Brun et al., 1997). By using long-slit spectroscopy it was then possible to confirm the redshifts of these objects (Fynbo et al., 2010, 2013; Rahmani et al., 2016). Supplementary to the wide-field search was the method of using a narrow-band optical filter (about 10 Å) centred around the expected [OII] emission at the absorbers redshift (Yanny et al., 1987, 1989, 1990a; Møller & Warren, 1993). Another detection method involved surveying the quasar field with a X-Shooter slit triangulation in search for associated galaxies (Møller et al., 2004; Krogager et al., 2017). Narrow filter integral field searches, using an array of optic fibers, for extended emission near QSOs with absorbers were also attempted (Yanny et al., 1990b).

Integral-field spectroscopy with instruments like SINFONI (Eisenhauer et al., 2003), MUSE (Bacon et al., 2006) or OSIRIS (Larkin et al., 2006), make it possible to obtain spectra of many sources in the field-of-view and efficiently classify galaxies associated with absorbers. Combined with high-resolution UV spectroscopy of quasars, these studies have revealed relevant features interpreted as galactic outflows (Schroetter et al., 2016; Rahmani et al., 2018b; Schroetter et al., 2019), warped accretion discs (Rahmani et al., 2018a) and galactic fountains (Bouché et al., 2016). Alongside the detection of kinematic traces of gas inflows and outflows, new studies detected multiple galaxies associated with one absorber (Bielby et al., 2017; Péroux et al., 2017; Klitsch et al., 2018; Nielsen et al., 2018; Péroux et al., 2019; Muzahid et al., 2020). These findings point towards a more complex view of the gaseous halos of galaxies: from the halo gas of single systems, through tidal streams and halo substructures (Whiting et al., 2006; Kacprzak et al., 2010) to intra-group gas (Bielby et al., 2017; Fossati et al., 2019). All these studies have however focused on single absorber systems or on single-species absorbers such as OVI (e.g. Bielby et al., 2019) or MgII (Zabl et al., 2019), lacking the information about the HI, which is crucial for metallicity estimates. At higher redshifts, some damped Lyα absorbers (DLAs, with $\log[N(\mathrm{HI})/\mathrm{cm}^{-2}] \geq 20.3$) and Lyman Limit Systems, (LLS, $\log[N(\mathrm{HI})/\mathrm{cm}^{-2}] \leq 19.0$) have been found to be associated with multiple Lyα Emitters (Mackenzie et al., 2019; Lofthouse et al., 2020). In this work, we undertake a statistical analysis of HI-selected absorbers. We present a MUSE study of five quasar fields with 14 HI absorbers at $z \leqslant 1.4$, to perform a statistical study of the properties of the absorbers and their associated galaxies, including their metallicities. Each of these fields also have associated ALMA observations, targeted to detect CO(2−1) and CO(3−2) emission lines from galaxies associated to HI rich quasar absorbers. Measuring the molecular content of these systems makes it possible to constrain the physical properties

and kinematics of molecular gas, known to be the fuel of star formation (Szakacs et al. in prep).

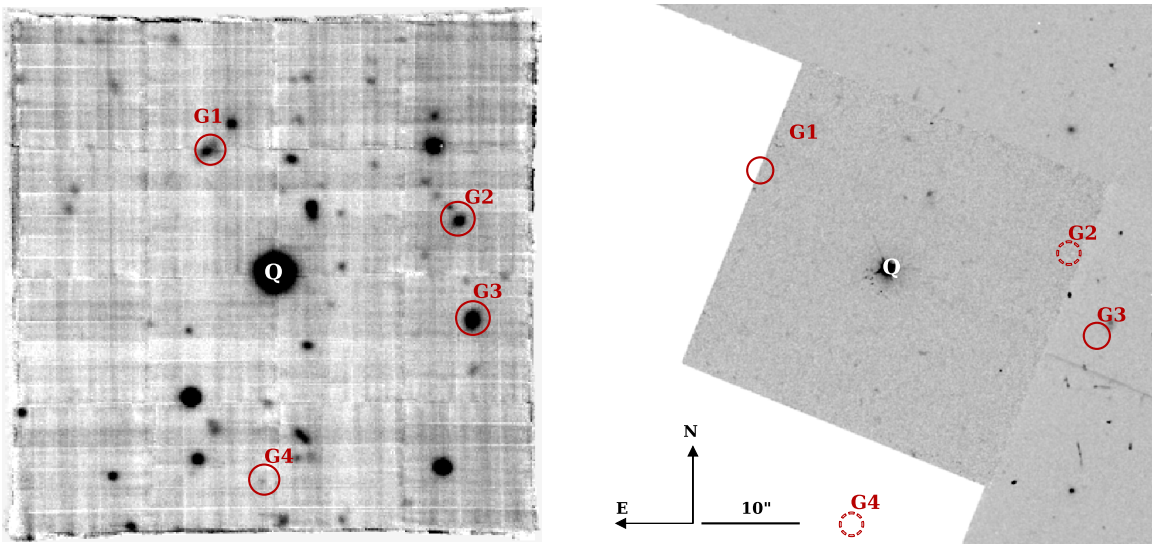

Figure 2.1: MUSE white light (left) and HST F702W (right) images of the Q1232z083 absorber field. Galaxies associated with this absorber are marked with red circles. Non-detections are marked with dashed circles: both galaxies G4 and G2 are undetected in the HST image, G4 because it falls outside the field of view of the WFPC2 camera, and G2 because it is at the edge between two camera chips. The MUSE and the HST images in the figure are both centred around the background Q1232-0224 ($z_{\rm QSO} = 1.05$). This figure illustrates the power of IFU observations to identify galaxies associated with absorbers. Images of the remaining fields from the sample with galaxies associated with the absorbers marked on them are presented in the Appendix A.

2.2 MUSE Observations and ancillary data

The MUSE-ALMA Halos sample of quasar absorbers is based on five quasar observed with MUSE and archival high-resolution UV-spectroscopy: Q0152-2001, Q1130-1449, Q1211+1030, Q1229-0207, Q2131-1207; quasars coordinates and redshifts are summarized in Table 2.1. Through the paper we refer to these quasars with their first four right ascension numbers (e.g. Q1130). Similarly, we adopt the naming scheme for the absorbers in our sample: first four right ascension numbers + z + the redshift of the absorber (up to second decimal point (e.g.Q1229z076).

The MUSE quasar fields were selected (proposal 096.A-0303(A), PI: Péroux) for the presence of known HI absorbers, with column density $\log[N(HI)/(\mathrm{cm})^{-2}] > 18.5$ at redshift z $\sim$ 0.4, for which MUSE covers all prominent galactic emission lines: Hβ, [OII], [OIII],

Hα, [NII], [SII]. Furthermore, for all selected quasars, at least one galaxy was known to be associated with this HI absorber, identified with photometry or low-resolution spectroscopy (Q0152 - Rao et al. (2011), Q1130 - Kacprzak et al. (2011), Q1211 - Kacprzak et al. (2011), Q1232 - Le Brun et al. (1997), Q2131 - Kacprzak et al. (2011)). For all targeted quasars absorbers we detect the known counterpart.

QSO name	alternative name	RA	DEC	z
	[hh:mm:ss]	[dd:mm:ss]		
Q0152-2001	UM 675	01:52:27	-20:01:07.1	2.06
Q1130-1449	B1127-145	11:30:07	-14:49:27.7	1.19
Q1211+1030	1209+107	12:11:41	+10:30:02.8	2.19
Q1232-0224	1229-021	12:32:00	-02:24:04.6	1.05
Q2131-1207	Q2128-123	21:31:35	-12:07:04.8	0.43

Table 2.1: Parameters of the quasars in MUSE-ALMA Halos sample: (1) reference name used in this paper (2) full name of the quasar, (3) right ascension, (4) declination, (5) quasar redshift

2.2.1 High resolution UV/optical quasar spectroscopy

We used UVES/VLT and HIRES/Keck archival high-resolution spectra for all five quasars in our sample to identify absorbers and their multiple metal absorption lines. Q0152-2001 and Q2131-1207 were observed with HIRES/Keck at resolution R = 45,000–48,000. The remaining quasar spectra (Q1130-1449, Q1211+1030 and Q1232-0224) we covered by the VLT/UVES spectrograph BLUE arm, with combined exposure times from 4 to 6 hours and spectral resolution up to R = 80,000. The wavelength coverage differs between objects due to different initial observation setups, introducing gaps in the spectra. The data were uniformly reprocessed using the ESO UVES pipeline and individual exposures of each target were merged, weighted by their signal-to-noise-ratios (Zafar et al., 2013).

We used also archival HST/FOS spectra covering the HI absorption line for all the absorbers in our sample. We used the already reduced FOS UV spectra from HST archives for Q1232-0224 (PI: Bergeron ID:5351, FOS-G190H, R = 1300, $T_{\rm exp}$ = 4580s) and Q1130-1449 (PI: Deharveng, ID: 3483, FOS-G160L, R = 250, $T_{\rm exp}$ = 1450s). For the remaining systems, we relied on the measurements of HI column density from the literature (references included in Table 2.2. We note that the wavelength reference frame differs for MUSE (air) and UVES/HIRES (vacuum). To obtain precise redshift and velocity measurements we shifted the MUSE spectra to a vacuum wavelength reference system.

2.2.2 MUSE and HST imaging of five quasar fields

The five quasar fields we present in this work were observed with VLT/MUSE IFU in period 96 (proposal 096.A-0303(A), PI: Péroux). All observations were conducted in good seeing conditions (< 0.85 arcsec) with an average of 1–2 hours per target for all but field Q1130-1449 (Péroux et al., 2019), which was significantly deeper (12 × 1200s). All fields were observed in nominal mode, resulting in a spectral coverage of 4800–9300 Å.

All raw MUSE observations of our quasar fields were reduced with the ESO default MUSE reduction pipeline v2.2 (Weilbacher et al., 2016). We applied bias, flat and wavelength calibration, and line spread functions as well as illumination correction frames to each individual exposure with *scibasic* command, creating MUSE specific PIXTABLE output. The astrometry solution and the correction for geometry, together with flux calibrations, were all combined within the *scipost* command. Finally, individual exposures were combined including field rotation. The detailed description of the reduction procedure can be found in Péroux et al. (2019). We refrained from using the pipeline sky subtraction method, due to unsatisfactory results. Instead, removal of sky emission lines was performed on the final datacube with a Principal Component Analysis (PCA) algorithm (Husemann et al., 2016). The code creates PCA component of selected sky regions used further to remove sky residuals.

In addition to MUSE observations we used readily available, reduced archival HST imaging for the galaxy identification: WFPC2 (Wide Field Planetary Camera 2) in filter F702W of quasars Q0152-2001 (PI: Steidel, ID:6557), Q1211+1030, Q1232-0224 (PI: Bergeron, ID:5351) and Q2131-1207 (PI: Maccheto, ID:5143) and WFC3 (Wide Field Camera) IR-F140W of Q1130-1449 (PI: Bielby, ID: 14594), with exposure times from 10 to 50 minutes. An example of MUSE and corresponding HST/WFPC2 quasar field images for Q1232z083 is shown in Figure 2.1.

2.2.3 MUSE data reduction

Abs ID	$\mathbf{z_{QSO}}$	$\mathbf{z_{abs}}$	$\mathbf{\log(N(HI)/cm^{-2})}$	W_r**(2796) [Å]**	**[Fe/H]**
Galaxy	z_{gal}	b [kpc/arcsec]	$SFR_{[OII]}$ [$M_\odot yr^{-1}$]	$12+\log(O/H)_{lo}$	$12+\log(O/H)_{up}$
Q0152z038	**2.06**	**0.3887**	**< 18.8** [f]	**0.58 ± 0.05** [a]	**> -1.36**
G1	0.3826	60 / 11.5	1.04 ± 0.03	8.65 ± 0.09	-
G2	0.3802	80 / 15.4	0.33 ± 0.01	8.70 ± 0.07	-
G3	0.3815	86 / 16.5	0.31 ± 0.01	8.42 ± 0.06	-
G4	0.3814	123 / 23.7	0.12 ± 0.02	8.94 ± 0.08	-
G5	0.3814	172 / 33.1	< 0.02	-	-
G6	0.3824	205 / 39.4	0.07 ± 0.01	7.82 ± 0.13	-
Q0152z078	**2.06**	**0.7800**	**18.87 ± 0.14** [d]	**0.36 ± 0.04** [d]	**> -0.44**

G1	0.7803	54 / 7.6	23.00 ± 5	7.90 ± 0.20	8.60 ± 0.20
Q1130z019	**1.19**	**0.1906**	**< 19.1**	**0.14 ± 0.01**	**-**
G1	0.1905	17 / 5.3	-	-	-
Q1130z031	**1.19**	**0.3127**	**21.71 ±0.07[b]**	**2.21 ± 0.12[d]**	**-1.94 ± 0.08 [g]**
G0	0.3131	11 / 2.3	0.07 ± 0.1	8.64 ± 0.14	-
G1	0.3121	18 / 3.8	2.80 ± 0.8	8.12 ± 0.08	-
G2	0.3127	44 / 9.5	0.44 ± 0.3	8.77 ± 0.05	-
G4	0.3126	82 / 17.7	> 0.40	< 8.65	-
G6	0.3115	98 / 21.3	1.14 ± 0.7	8.94 ± 0.16	-
G16	0.3133	21 / 4.5	0.42 ± 0.1	7.96 ± 0.14	-
G17	0.3136	35 / 7.6	3.18 ± 0.7	8.01 ± 0.17	-
G18	0.3125	27 / 5.9	0.30 ± 0.1	8.69 ± 0.18	-
G19	0.3119	120 / 26.0	1.09 ± 0.2	8.67 ± 0.16	-
G20	0.3140	56 / 12.1	< 0.01	-	-
G21	0.3122	131 / 28.4	0.01 ± 0.01	8.76 ± 0.17	-
Q1130z032	**1.19**	**0.3283**	**< 18.9**	**0.028 ± 0.003 [a]**	**-**
G3	0.3282	74 / 15.3	0.02 ± 0.01	7.52 ± 0.28	8.93 ± 0.09
G5	0.3284	87 / 18.2	0.006 ± 0.001	-	-
Q1211z039	**2.19**	**0.3929**	**19.46 ± 0.08 [d]**	**1.19 ± 0.01 [a]**	**> -1.05**
G1	0.3928	37 / 6.8	4.71 ± 0.08	8.16 ± 0.01	8.48 ± 0.01
Q1211z062	**2.19**	**0.6296**	**20.30±-0.30 [d]**	**2.92 ± 0.23 [d]**	**-0.98 ± 0.3 [e]**
G1	0.6283	12 / 1.7	1.04 ± 0.14	-	-
G2	0.6303	132 / 19.1	0.59 ± 0.05	-	-
Q1211z089	**2.19**	**0.8999**	**< 18.50**	**0.023 ± 0.005**	**-**
G1	0.8991	79 / 10.0	4.33 ± 0.43	-	-
G2	0.8995	185 / 23.4	2.03 ± 0.19	-	-
G3	0.8991	275 / 34.8	0.57 ± 0.07	-	-
G4	0.8953	149 / 18.9	3.29 ± 0.14	-	-
G5	0.8960	197 / 24.9	1.68 ± 0.21	-	-
Q1211z105	**2.19**	**1.0496**	**< 18.90**	**0.18 ± 0.01**	**> -1.69**
No galaxies found to match the absorber					
Q1232z039	**1.05**	**0.3950**	**20.75 ± 0.07 [e]**	**2.894 ± 0.11**	**< -1.31 [e]**
G1	0.3953	8 / 1.5	0.67 ± 0.09	8.02 ± 0.06	8.66 ± 0.04
Q1232z075	**1.05**	**0.7572**	**18.36 ± 0.09 [d]**	**0.52 ± 0.07 [d]**	**> -1.48**
G1	0.7566	68 / 9.1	2.58 ± 0.23	8.54 ± 0.19	8.19 ± 0.19

Q1232z076	**1.05**	**0.7691**	**18.11 ± 0.15**	**0.053 ± 0.001**	**> -2.34**
G1	0.7666	57 / 7.7	3.85 ± 0.14	-	-
G2	0.7685	131 / 17.5	0.84 ± 0.07	8.06 ± 0.14	8.55 ± 0.10
G3	0.7688	170 / 22.7	<1.41	-	-
Q1232z083	**1.05**	**0.8311**	**18.84 ± 0.10**	**0.238 ± 0.004**	**> -2.19**
G1	0.8309	122 / 15.8	3.10 ± 0.25	-	-
G2	0.8323	166 / 21.6	< 0.70	-	-
G3	0.8325	176 / 22.8	5.42 ± 0.29	-	-
G4	0.8321	182 / 23.6	0.61 ± 0.13	-	-
Q2131z043	**0.43**	**0.4298**	**19.50 ± 0.15** [c]	**0.41 ± 0.01** [d]	**> -0.96**
G1	0.4301	52 / 9.2	2.00 ± 0.2	8.98 ± 0.02	-
G2	0.4307	61 / 10.7	0.20 ± 0.1	8.32 ± 0.16	-
G3	0.4301	147 / 26.0	0.08 ± 0.1	-	-
G4	0.4298	174 / 30.7	0.10 ± 0.1	-	-

Table 2.2: Parameters of the absorbers and their associated galaxies. Each absorber row is followed by the rows of associated galaxies. Absorber parameters: (1) AbsorberID: first part of the quasar name + the absorber redshift, selection method in brackets (2) redshift of the quasar, (3) redshift of the absorber, (4) column density of HI, (5) Equivalent width of MgII λ2796, (6) absorber [Fe/H] (not dust corrected). Each galaxy parameters row consists of: (1) Galaxy ID, in the order of increasing impact parameter, unless the nomenclature adapted from literature states differently (Q1130z031, Q1130z032), (2) redshift of the galaxy, (3) impact parameter in kpc and arcsec, (4) star formation rate measured from [OII] emission line flux (not dust corrected) Kobulnicky et al. (1999), (5–6) 12 + log(O/H) (not dust corrected): we report both metallicity branches following Kobulnicky et al. (1999); in case of literature values where only one metallicity was reported we leave the upper metallicity column blank.
Literature references: a - Kacprzak et al. (2011), b - Lane et al. (1998), c - Muzahid et al. (2016), d - Rao et al. (2006),e - Boisse et al. (1998), f - Rahmani et al. (2018b), g - Péroux et al. (2019)

2.3 Analysis and Results

2.3.1 Identification of HI and metal absorbers in high-resolution UV quasar spectra

High-resolution UV quasar spectra are used for identification of the metal absorbers and measurements of the metal line column densities. Together with far-UV measurements of HI column density, this allowed us to measure metallicities of the observed gas. The primary sample is comprised of known strong HI absorbers, we used the high-resolution UV spectra to identify the metal lines associated with these absorbers. Specifically, we identified FeII (FeII λ 2382, 2344, 2586, 2600), which we use for the metallicity measurements. We then expanded the sample to other metal absorbers by identification of the characteristic lines such as MgII or CIV doublets down to the equivalent width limit of

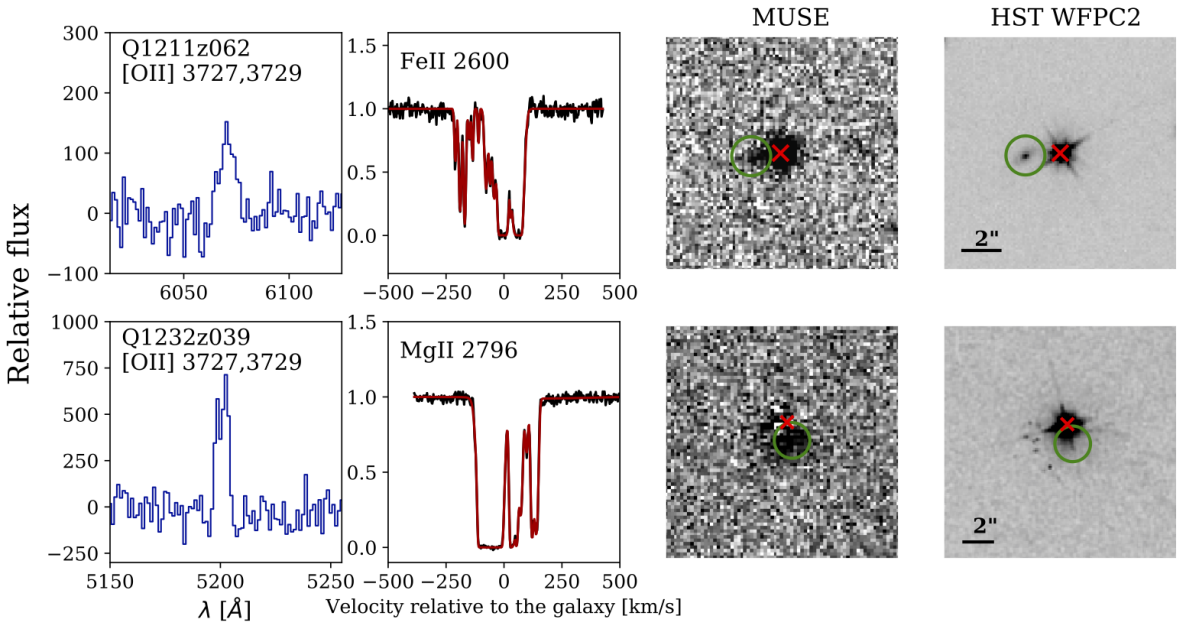

Figure 2.2: Spectral quasar PSF subtraction in the MUSE data. The panels present two galaxies each hidden under the PSF of a different QSO, and their retrieved spectra: Q1211+1030–top and Q1232-0224–bottom. Galaxies were identified by detecting the [OII] emission line (first panel from the left) at the redshift of the absorber. To remove the quasar signature we subtracted the continuum around the detected line. The second column presents the absorption lines detected in the normalized quasar spectrum in velocity space with zero velocity set to the systematic redshift of the galaxy. The third and fourth columns show the continuum-subtracted zoom-in MUSE image of the galaxy around the [OII] emission line and the zoom-in of the HST F702W image of the quasar on the same scale. The red cross represents the centre of the quasar and green circle the position of the detected galaxy.

$W_{rmr}(2796) > 0.02$. For the redshift determination, we used MgII λ2796; in case of multiple components, we used the strongest one to measure the central wavelength. We limited the sample to systems with redshift $z < 1.4$, so that the [OII] emission line is within the MUSE wavelength coverage, resulting in a list of 14 MgII absorbers, of which some have been presented elsewhere: Q0152z038 (Rahmani et al., 2018a), Q0152z078 (Rahmani et al., 2018b), Q1130z031 (Péroux et al., 2019) and Q2131z042 (Péroux et al., 2017). We used VPFIT [1]v10.0 to fit the Voigt profiles to unsaturated metal absorption lines. In particular, we fitted FeII of systems whose metallicities were not known from the literature. We referred to the measurements of HI column densities (N(HI)) in the literature, and for missing cases (Q1211z089, Q1211z105, Q1232z076, Q1232z083) we measured it used the archival HST FOS UV spectra. We derived [Fe/H] metallicities of the absorbing gas

[1]https://www.ast.cam.ac.uk/∼rfc/vpfit.html

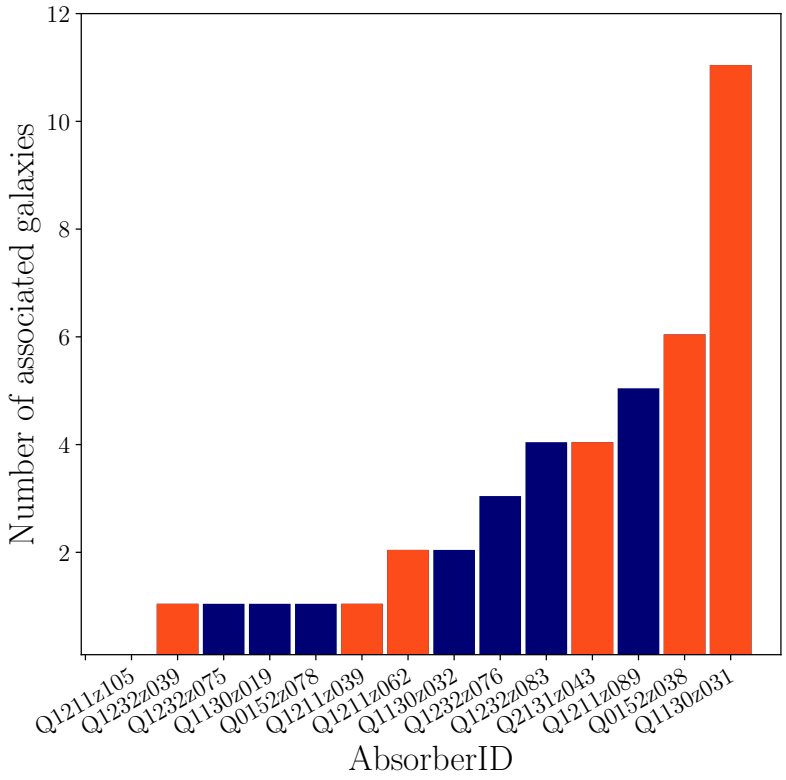

Figure 2.3: Number of galaxies associated with each absorber in our sample. HI selected galaxies are indicated in orange, metal-selected ones in blue. Most of the absorbers have two or more galaxies at the associated redshifts. Q1130z031 is an exceptionally rich galaxy group and also has the highest HI column density (Péroux et al., 2019). For Q1211z105 we do not detect any galaxy at the redshift of the absorber down to 0.1 $M_{\odot}yr^{-1}$.

following the equation:

$$[Fe/H] = log_{10}\left(\frac{N_{Fe}}{N_H}\right) - [Fe/H]_{\odot}, \tag{2.1}$$

(where N_x is the column density of an ion x) and using a solar metallicity value of $[Fe/H]_{\odot}$ = -4.5 from Asplund et al. (2009). However, Fe is not a perfect tracer of the metallicity since it is well known that this element can be depleted onto dust.

Dust in the gas associated with the absorber could cause reddening of the quasar spectrum. To test if there is a significant amount of dust associated with the absorbing gas, we fit the spectral templates to quasars optical spectra extracted from MUSE cubes and archival optical and near-infrared photometry. The details of the method can be found in the Appendix A. We estimate the reddening to be $E(B-V)$ = 0.02 - 0.03 for all quasars in the sample, indicating that the that Fe depletion onto dust in absorbers is likely small. Given the remaining uncertainty in the dust correction, we considerably refer to absorption metallicity estimates as lower limits.

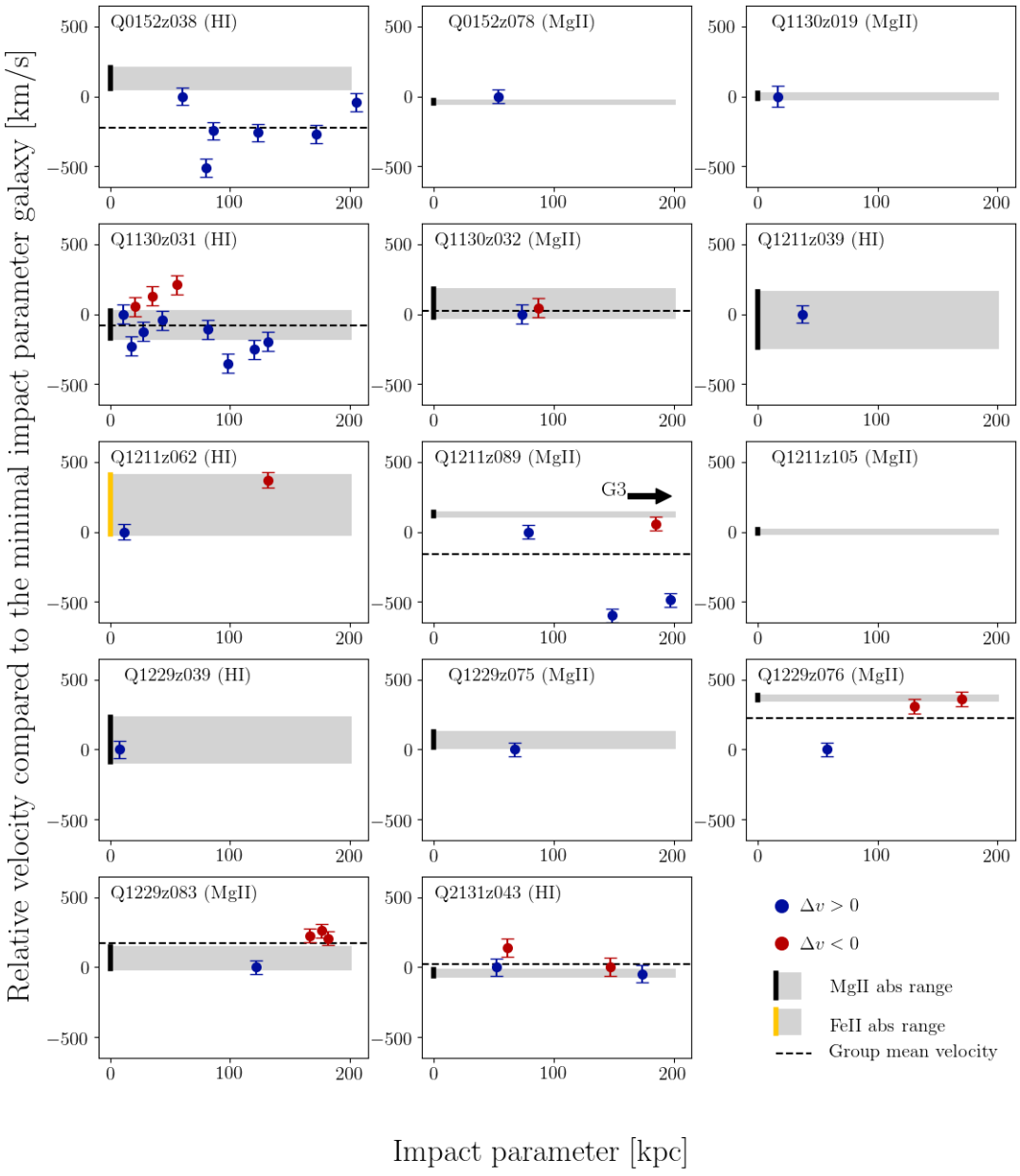

Figure 2.4: Impact parameter - velocity diagram of the MUSE-ALMA Halos sample. The velocity of the galaxy at the lowest impact parameter is set to zero and the velocities of the remaining galaxies and absorption lines are calculated with respect to that galaxy. Galaxies with Δv > 0 are marked in red and with Δv ≤ 0 in blue. The black line (and the further grey area) indicates the velocity span of all MgII λ2796 absorption line components. For the Q1211z062 system (marked in yellow) MgII λ2796 was not covered and FeII λ2600 was used instead. Dashed black like marks the mean velocity of the galaxy group. Note that galaxies with the smallest impact parameters are not always closest to the gas in velocity.

2.3.2 Ionization correction for low N(HI) absorbers from Bayesian MCMC approach

Apart from the dust, the absorption metallicity measurement can be affected by the H ionization fraction. Most absorbers in our sample are not DLAs ($\log(N(\mathrm{HI})) < 20.3$ cm^{-2}) and in such systems we expect a significant portion of the hydrogen to be ionized, influencing the metallicity derivations. To assess the effect of the ionisation correction on the metallicities derived for such absorbers, we used the radiative transfer calculations from CLOUDY (Ferland et al., 2013). We performed Bayesian MCMC modelling on the pre-computed grid of Cloudy models from Fumagalli et al. (2016). The grid consisting of multiple radiative transfer models is defined by four parameters: N(HI) ($\log(N(\mathrm{HI})) =$ 17.0 - 20.5), redshift (z = 0.0 - 4.5), metallicity ($\log(Z/Z_{\odot}$ = -4.0 - 1.0 and n_H density ($\log(n_H)$ = -4.5 - 0.0). N(HI) and redshift are constrained from the observations and serve as an input. To determine the most probable model, we use as constraints the measured metal line column densities (MgII, MgI and FeII in almost all cases). We run the MCMC chain over the grid, using the PyIGM package routines [2] (Prochaska et al., 2017). For the resulting Probability Density Functions (PDF) of each of the parameters, we adopt the median value of the computed [Fe/H] metallicity (see Table 2.2). For a detailed description of the method, its extension and assumptions in the UV background we refer the reader to Fumagalli et al. (2016) and Wotta et al. (2019).

2.3.3 Identification of galaxies associated with absorbers

After identification of all quasar absorbers, we used the MUSE observations of the quasar fields, together with wide-field imaging with HST/WFPC2 to search for galaxies associated with these absorbers. We used several analysis techniques to maximize the completeness of the search, including different types of objects (SF-galaxies, passive galaxies, faint, low SFR objects etc.). MUSE datacubes can be viewed as individual narrow-band (NB) images at each wavelength slice or a combined continuum image over the whole observed wavelength range, allowing for two types of object searches in the data: a single spectral-line search at the redshift of the absorber and identification of the sources seen in continuum. We combined 50 (50$\times$ 1.25 Å) MUSE wavelength planes into a pseudo-narrow-band (NB) images centred at the expected [OII] observed wavelength at the redshift of the absorbers. The choice of the width of the pseudo-NB corresponds to a velocity difference to the absorber redshift of $\Delta v \pm$ 1000-1500 km s^{-1}, depending on the central wavelength of the observed [OII] line. The slight redshift dependence is a mild effect which is mitigated by the systemic search with MUSELET. Then, we identified all bright sources on the NB images and inspected their full spectra to confirm the redshift by the presence of other emission lines (if detected). This allows us to detect galaxies at a range of impact parameters, from systems close to the quasar ($<$ 1 arcsec) up to the limits imposed by the MUSE field-of-view. Within the spectral coverage of MUSE, [OII] doublet can be observed up to z =

[2] PyIGM Package is available at `https://github.com/pyigm/pyigm`

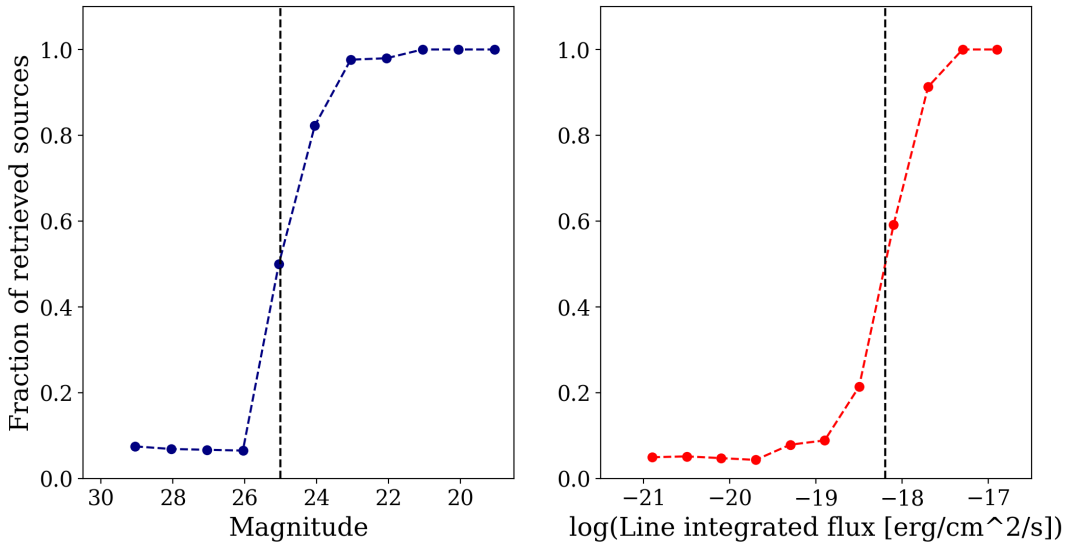

Figure 2.5: Completeness of the MUSELET search for continuum sources and line emitters in MUSE-ALMA Halos MUSE cubes. Detection fraction of line emitters with continuum as a function of their magnitude are on the top panel (blue line) and line emitter without detected continuum as a function of a integrated line flux are on the bottom panel (red line). The black dashed line marks the 50 percent completeness level.

1.4, however from $z = 0.8$ it will be the only prominent galactic line covered. We exclude the possibility of the contamination from the low-z sources ($z < 0.8$), because at that redshift ranges we would see more then one line (bright [OIII] or Hα). For the higher-z case ($z > 1.4$), the only observable emission line is Lyα (covered by MUSE at $z = 2.7$ - 6.7). Although theoretically, the detection could be a Lyα within that redshift range, based on the LAE luminosity function and lack of the expected prominent asymmetric shape we exclude that interpretation.

We search for galaxies within $\Delta v\pm$ 1000 km s^{-1} from the absorber. We measure the centre of the lines with a precision below 1 Å
(corresponding to an error in the redshift estimation smaller than $\Delta z \pm 0.0003$). We note that the majority of galaxies lie within $\Delta v\pm$ 500 km s^{-1}.

For three absorbers, Q1211z062, Q1232z039 and Q1130z031 (Péroux et al., 2019), we detect galaxies close to the quasar position (within 1 arcsec), blended with the quasar PSF. Spectral PSF subtraction allows for easier detection of such systems, despite the quasar bright contribution. In each case, we detected the [OII] emission doublet at the redshift of the absorber. We extracted the 1D spectrum of 4 $\times$ 4 spatial pixels around the line detection. The continuum of these objects are not detected. The method is illustrated in Figure 2.2, for absorbers from two of the QSOs, where we present a zoom into the extracted spectrum around the [OII] emission line (in the observed wavelength frame), the

main absorption feature of the associated absorber in the galaxy velocity space as well the pseudo-narrow band image alongside the HST image of the same region.

After narrow-band identification of the sources at the redshift of the absorber, we extracted all continuum sources from the MUSE white light image using the MUSELET source finder, part of the MPDAF package provided by the MUSE GTO team (Bacon et al., 2016). MUSELET creates from a cube: (i) a white light image; (ii) R, G, B images corresponding to 1/3 of the wavelength range each; (iii) a set of narrow-band images based on the average of 5 wavelength planes (total width of 6.25 Å). It then runs SExtractor (Bertin & Arnouts, 1996) on these products, resulting in a catalogue of continuum sources. We classified all of the reported sources as either emission line galaxies, absorption line galaxies, stars, unknown continuum sources (without emission or absorption lines) or artifacts. We inspected all extracted spectra and used [OII] and [OIII] emission lines or Ca K+H absorption lines to determine the redshift of each source. Among all galaxies associated with absorbers, we detected four passive galaxies (with only Ca K+H absorption lines and red continuum). For completeness, we cross-matched the catalogue of continuum sources detected in MUSE with detections in archival HST/WFPC2 images (Le Brun et al., 1997). This search led to no new galaxies discovered at the redshift of the absorbers. After identification of the galaxies associated with the absorber on the white light continuum MUSE image, we measured their position with respect to the quasar. We fitted a 2D Gaussian to the continuum images of both the quasar and the galaxies to determine their geometrical centres and then measure the angular distance between them.

To test the completeness of the methods described above, we injected 100 mock sources into each MUSE datacube. The mock sources are single Gaussian emission lines of the integrated flux ranging from 10^{-21} to 10^{-17} erg s^{-1} cm^{-2}, represented spatially by a 2D Gaussian of FWHM of 3 pixels, corresponding to the MUSE PSF. We introduce the two types of mock sources: line emitters with continuum flux and without continuum. After running the MUSELET source extraction on these cubes we compared the catalogue of detected sources with the list of mock sources. The completeness of our search is presented in Figure 2.5. We are over 90 percent complete for detections brighter then 23.5 mag for continuum objects and line flux greater then $10^{-17.7}$erg cm^{-2} s^{-1}. We can retrieve sources down to the magnitude of 25.7 mag and flux $10^{-18.5}$erg cm^{-2} s^{-1}. We detected 43 associated galaxies within a 250 kpc impact parameter from the QSO sight-line. As we summarize in Figure 2.3, we detect systems with multiple galaxies associated with the absorber as well as a few systems where only one galaxy was detected. Only in one case we do not detect any associated galaxy. This results in a high detection rate of 93 per cent. Note that five of the absorbers in the sample were targeted MUSE observations of absorbers with a known counterpart. Nevertheless, among the remaining nine absorbers, we detect at least one galaxy for eight absorbers. Altogether, among the 14 absorbers, eight (57 per cent) have more than one associated galaxy. This high detection rate is likely related to combination of the wide field of view and simultaneous spectral coverage of MUSE. We are probing a large range of impact parameters, from low (< 10 kpc) up to large distances of 200-250 kpc. We detect emission line sources and passive galaxies with no emission lines but solely detected in continuum at the redshift of the absorbers down to an integrated line flux of

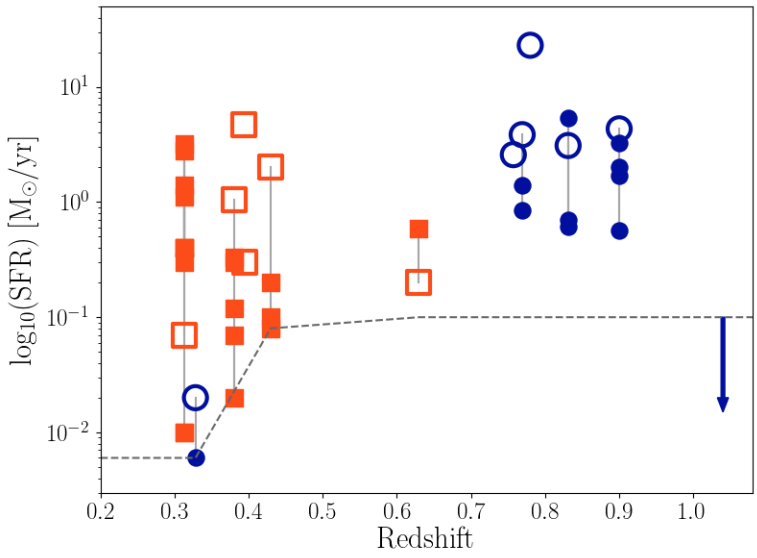

Figure 2.6: The star formation rate ($SFR_{[OII]}$) of the galaxies associated with absorbers as a function of redshift. Vertical grey lines connect galaxies associated with the same absorber. Open symbols indicate the galaxy at the smallest impact parameter. Absorbers selected based on their HI absorber are marked with orange squares, while metal-selected ones are marked with blue circles. The arrow marks the limits of the non-detection of the galaxy associated with the absorber Q1211z105. The dashed grey line represents the SFR limit at the redshift of the absorbers. We reach star formation rates of 0.1 $M_\odot yr^{-1}$ for most of the sample and lower (about 0.01 $M_\odot yr^{-1}$) in case of deeper data (Q1130-1449).

$10^{-18.5}$erg cm^{-2} s^{-1} and magnitude of 25.7 mag at the 20 percent completeness level.

Table 2.2 summarizes the parameters of all absorbers and their associated galaxies. Spectra of all systems alongside their position marked on MUSE white light images and HST images are shown on the Figure 2.1 and in the Appendix A.

2.3.4 Measurements of the physical properties of galaxies associated with absorbers

The majority of the galaxies associated with absorbers in our sample are emission-line objects. We can use the measured line fluxes to calculate the star formation rates of these galaxies as well as their emission-line metallicities. To obtain line fluxes, we fit Gaussian profiles to the [OII], [OIII], Hβ and Hα emission lines (if detected), using the MPDAF Gaussian profile fitting tool. For the redshift estimation, we use [OII]$\lambda\lambda$3727, 3729 (fitted with double Gaussian profiles) as this line is detected in all galaxies. In the case of galaxies lacking emission lines, we used Ca K+H absorption lines to determine the redshift.

For each galaxy we compared the velocity offsets between the members of the group and the absorbing gas. In Figure 2.4 we present the relative velocities of the galaxies of

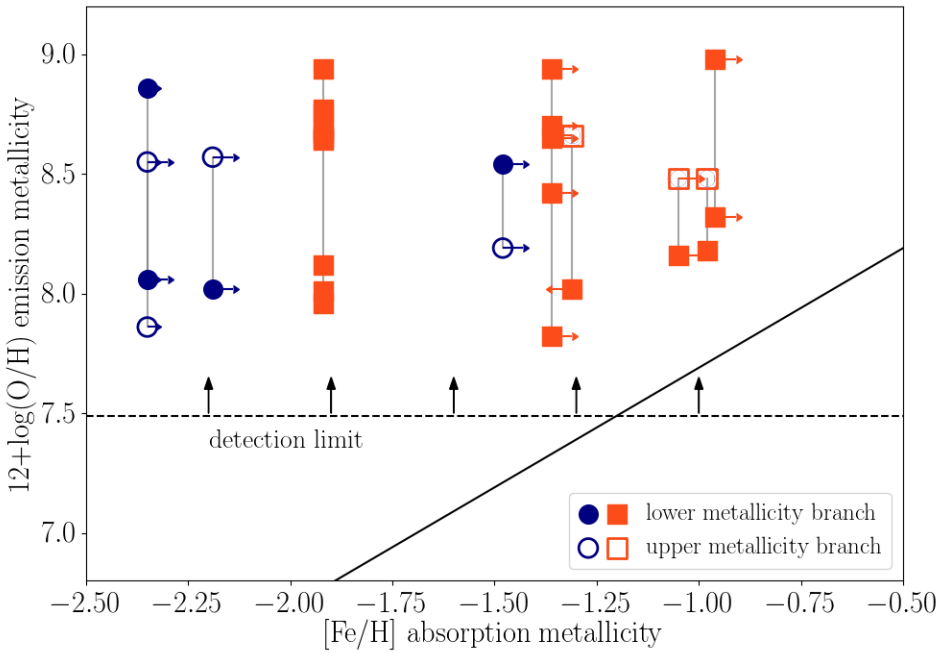

Figure 2.7: Oxygen abundance (emission metallicity) of the galaxies versus the metallicity of the absorbing gas. The solid black line represents the x=y line, the grey dashed line marks the detection limit of our data. Galaxies associated with the same absorber are linked with a vertical line. Filled and hollow circles represent upper and lower metallicity branch solution. Absorbers selected based on their HI absorber are marked with orange squares, while metal-selected ones are marked with blue circles. Measurements from the literature with one solution are also plotted. In all cases where it is measured, the galaxy emission metallicity is larger than that of the absorbing gas.

each absorbing system calculated with respect to the galaxy at the lowest impact parameter. Therefore, the velocity of the galaxy at the lowest impact parameter is set to zero. Additionally, we calculated the velocity between that galaxy and the absorbing gas. In four cases (out of eight multi-galaxy systems) objects other than the closest galaxy are better aligned in velocity with the gas seen in the absorption.

For all emission-line galaxies in the sample (except Q1130z019, for which we used Hα) the [OII] emission line is covered and detected. We used the flux of the line to calculate the star formation rates of the galaxies following the equation 2.2 from Kennicutt (1992).

$$SFR_{[\mathrm{OII}]} = 1.4 \times 10^{-41} \times L_{\mathrm{OII}}[M_{\odot}/yr] \tag{2.2}$$

In Figure 2.6 we present the measured SFR as a function of the redshifts of the absorbers. For sources for which we do not detect emission lines, we calculate the 3σ line limits at the expected line positions.

We detect sources at the redshift of the absorbers down to an SFR of 0.01 $\mathrm{M_{\odot}yr^{-1}}$ for $z < 0.4$ and 0.1 $\mathrm{M_{\odot}yr^{-1}}$ for higher redshifts (Figure 2.6). We report galaxies with a range

of SFR from passive galaxies (< 0.1 $M_{\odot}yr^{-1}$) to star-forming galaxies. Except for one case (23 $M_{\odot}yr^{-1}$) the SFR of associated galaxies range from 1 to 10 $M_{\odot}yr^{-1}$). These values are not dust-corrected since for objects most Hα lays outside the probed wavelength range, thus making measurement of Balmer Decrement, the reddening proxy, impossible. Note that the parameters taken from literature also do not have any dust-correction applied.

To obtain the galaxy metallicities we chose to use only those galaxies with measurements of the oxygen species [OII], [OIII] and Hβ. We then used the R23 oxygen abundance calculation from Kobulnicky et al. (1999) as the metallicity proxy. This procedure results in two solutions lower (Equation 2.3) and upper metallicity branch (Equation 2.5):

$$OH_{low} = 12 - 4.944 + 0.767 \times x + 0.602 \times x^2 - y \times (0.29 + 0.332 \times x - 0.331 \times x^2) \quad (2.3)$$

where,

$$x = log\left(\frac{[\mathrm{OII}]\lambda 3727 + [\mathrm{OIII}]\lambda\lambda 4959, 5008}{H\beta}\right) \quad (2.4)$$

$$\begin{aligned} OH_{upp} = 12 - 2.939 - 0.2 \times x - 0.237 \times x^2 - 0.305 \times x^3 - 0.0283 \times x^4 + \\ - y \times (0.0047 - 0.0221 \times x - 0.102 \times x^2 - 0.0817 \times x^3 - 0.00717 \times x^4) \end{aligned} \quad (2.5)$$

where,

$$y = log\left(\frac{[\mathrm{OIII}]\lambda\lambda 4959, 5008}{[\mathrm{OII}]\lambda 3727}\right) \quad (2.6)$$

Ideally, additional parameters based on [NII] and Hα can be used to differentiate between the branches. However, these lines are not covered with MUSE and therefore we present both solutions. For passive galaxies, we use the line detection limits. For Q1211z089 and Q1232z083, the 12+log(O/H) values could not be determined because both [OIII] and Hβ are outside of the MUSE wavelength coverage.

A comparison of the metallicity of the galaxies with that of the absorbing gas is shown in Figure 2.7. Groups of galaxies associated with single absorbers are connected with vertical lines, and for most of the systems, the two metallicity branches are plotted. We mark the one-to-one relation with a black thick line in the lower part of the diagram. The red dashed line in Figure 2.7 marks the emission line metallicity limit of our measurements. For all measured systems, the galaxy metallicity is found to be higher than that of the absorbing gas although given the limits, the metallicities still could be consistent with the 1:1 relation. Note, however, that given the metallicity estimate caveats these result must be taken with some caution. Finally, since we are not sensitive to metallicities below 12+log(O/H) = 7.5, we might be missing faint low-metallicity systems in our sample of galaxies associated with absorbers.

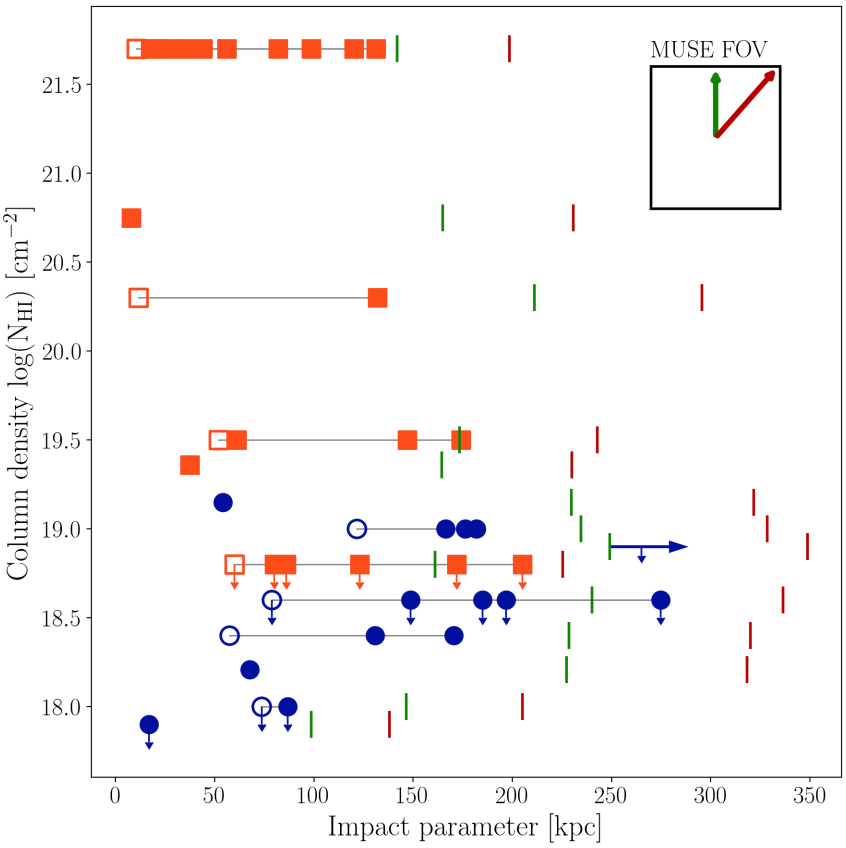

Figure 2.8: HI column density as a function of impact parameter. Absorbers selected based on their HI feature are marked with orange squares, while metal-selected ones are marked with blue circles. The blue, horizontal arrow marks the limits of the Q1211z105 associated galaxy impact parameter non-detection. The hollow symbols represent the lowest impact parameter object in groups. Red and green lines mark the limit of the impact parameter at the redshift of the absorber as defined by the MUSE field of view. Because the MUSE field of view is a square, we can reach larger impact parameters on the diagonal (red line) than along the square's side (green line; see the inset in the right upper corner). Horizontal grey lines connect galaxies associated with a single absorber. We find associated galaxies up to the impact parameters of 250 kpc.

2.4 Discussion

2.4.1 Absorbers are associated with multiple galaxies

The MUSE IFU proves to be an excellent instrument to search for galaxies associated with absorbers. The MUSE field of view of 1 arcminute translates to 200-250 kpc at redshift $z \sim$1. Since each object comes with a spectrum, we can immediately classify all objects in the field of view, including passive systems, increasing the extent of the search.

Furthermore, identification of low impact parameter galaxies ($<$ 1 arcsec) is also possible. Low impact parameter sources are the primary candidates for the absorbers hosts but can be potentially hidden under the extended PSF of bright quasars. Using spectral PSF subtraction, we can remove the quasar contamination and detect the emission lines from the source hidden under the PSF of the quasar (reference to Section 3.2). We report three such cases in our sample: Q1211z062, Q1232z039 (Figure 2.2) and Q1130z031 (Péroux et al., 2019).

Candidates for galaxies associated with absorbers were traditionally identified on wide-field images and later confirmed using long-slit spectroscopy. A number of presumably unique galaxy-absorber pairs at different redshifts were discovered this way, supporting the picture of a CGM tracing the halo of a single galaxy. Wide-field IFU spectroscopy improved the completeness of the absorber counterpart search, resulting in discoveries of multiple galaxies associated with absorbers (Bielby et al., 2017; Klitsch et al., 2018; Bielby et al., 2019; Fossati et al., 2019). These findings also extend to higher redshifts with $z \sim 3$ DLAs associated with multiple Ly α Emitters (LAE), as in (Mackenzie et al., 2019). Discoveries of multiple galaxies associated with low-z absorbers challenge the classical view of the CGM connected to the single halo of an isolated galaxy.

Cuts in both impact parameter and velocity difference between absorber and galaxy differ between published MUSE absorber studies, making the comparison of results not straightforward. MusE GAs FLOw and WInd (MEGAFLOW), a MUSE survey of MgII absorbers, (Schroetter et al., 2016) allows for a 1000 km s^{-1} velocity difference between the galaxy and the absorber, and in case of multiple detections, the lowest-impact-parameter object is selected. Quasar Sightline And Galaxy Evolution (QSAGE), an OVI absorption oriented MUSE program (Bielby et al., 2019), supported by a larger impact parameter coverage with HST grisms, uses a tight limit of 400 km s^{-1} (expected outflow velocity). For sources covered only by HST, the velocity range is extended to 700 km s^{-1} to account for the lower resolution of the HST grism spectra compared to MUSE. In a targeted single-system study, Bielby et al. (2017) find a galaxy group tightly aligned within $\Delta v = 140$ km s^{-1} around the absorber redshift, and within an impact parameter of 300 kpc, while in the case of a CO pre-selected absorber, Klitsch et al. (2018) looked for counterparts within $\pm$ 2000 km s^{-1}. In MUSE-ALMA Halos, we select galaxies to be associated with the absorber if their velocities does not differ by more than 1000 km s^{-1} from the absorber. Impact parameters are only limited by the MUSE field of view (up to 250 kpc at $z \sim$1). We mostly detect systems with multiple galaxies associated with the absorber in velocity range between few km s^{-1} to $\pm$ 750 km s^{-1}. We used the luminosity function of [OII]

emitters from Ly et al. (2007) to calculate the expected number of galaxies in the volume at the redshift of the absorbers. We find the expected number of detectable sources to be 0.45 for the higher redshift absorber and around 0.1 for the rest of the sample. Therefore we conclude, that we detect over-densities of galaxies associated with quasar absorbers, compared to random field galaxies.

Groups of galaxies typically include between 3–30 members and have a total dynamical mass of $(10^{12.5}$–$10^{14})h^{-1}$ $M_\odot$ with sizes between $(0.1$–$1)h^{-1}$ Mpc (Bahcall, 1996). Depending on mass, we would expect a velocity dispersion of group members between 100 and 500 km s^{-1} (Bahcall, 1996). For the groups of absorption-selected galaxies in our sample, the velocity dispersion spans 160–350 km s^{-1}. Such values are typical of groups environment.

Detecting multiple galaxies associated with the absorber introduces a complexity to the CGM–galaxy identification. Typically, absorbers at the lowest impact parameter would be considered the most likely counterpart for the absorber. Our analysis clearly demonstrates that impact parameters alone are not conclusive for identifying a unique galaxy associated with the quasar absorber. In Figure 2.4 we present the velocity - impact parameter diagrams for all absorbers in our sample. In several cases, the galaxy at the lowest impact parameter is offset from the gas velocity by more than the typical galaxy rotation velocity (150–200 km s^{-1}). By contrast, some galaxies with larger impact parameter are better aligned to the absorbers velocity. This does not point directly to the gas origin since we expect also to see tracers of outflows in the CGM and wind velocities can be as high as 600 kms^{-1}. To link the absorbing gas with a particular galaxy, a detailed component-by-component analysis of the absorbing gas profile alongside resolved kinematic studies of the galaxy are needed.

To put our sample in the context of other known absorbers, we test if the galaxies follow the canonical relations. In Figure 2.8 we present the HI column density versus impact parameter. Studies so far, mostly for high HI column density absorbers (Krogager et al., 2017; Augustin et al., 2018), showed an anti-correlation between N(HI) and impact parameter, which is also reproduced by the OWLS hydro-dynamical simulations (Rahmati & Schaye, 2014): the larger the N(HI) column densities the smaller the impact parameter. Including all galaxies associated with absorbers, we are retrieving this relation also at lower HI column densities (from 10^{18} to $10^{21.7}$ cm^{-2}). Multiple associated galaxies, as in our sample, introduce a significant scatter into this relation. The relation should be then seen as an "envelope" of possible values of N(HI) and impact parameters for associated galaxies.

All HI absorbers in our sample are also MgII absorbers. Having both Lyα and MgII for all systems is a unique quality of the sample. Typically, MgII absorbers follow an $W_r(2796)$ (MgII λ2796 equivalent width) and impact parameter anti-correlation (e.g. Lanzetta & Bowen, 1990; Bergeron & Boissé, 1991; Bouché et al., 2006; Chen et al., 2010). We compared our groups to several other MUSE studies from the literature and the MgII absorbers compilation MAGIICAT Nielsen et al. (2013) of isolated galaxies as well as galaxy groups (Nielsen et al., 2018). Figure 2.9 shows that MUSE-ALMA Halos systems are in agreement with this relation, especially with other absorption-related groups. Together with the group compilation from MAGIICAT, our data points seem to be offset from the indi-

vidual galaxies. Similar findings were reported by Bordoloi et al. (2011) in the COS-Halos comparison of MgII absorbers in groups and isolated galaxies. The $W_r(2796)$–b relation for multiple associated galaxies define an envelope of possible parameters for the absorbers. Despite being sensitive to both low MgII equivalent widths and low impact parameters, our observations do not populate the part of the plot with low $W_r(2786)$ and low impact parameters.

Lowest impact parameter systems can lie far from the absorbing gas in velocity space introducing the question of the nature of the gas detected in absorption and its physical connection the galaxies detected in emission. Detailed component-by-component analysis of the metal absorption lines, together with resolved kinematic and metallicity maps of the galaxies will likely provide information to identify the origin of the absorbing gas. Intra-group gas can be the result of the interactions between the group members, with possible higher density tidal streams or be filled with clouds of material infalling onto the centre of the group dark matter halo. This intra-group medium proves to be more complicated to describe and understand than a single halo picture, and is less straightforward to connect the inflow/outflow features with particular members of the group.

2.4.2 Metallicities of absorbers and associated galaxies

Measurements of HI and metal column densities of the absorbers in our sample give a unique opportunity to compare directly the metallicity of the gas and the integrated metallicity of the associated galaxies. Our absorption metallicity measurements suffer from several caveats which can affect the results presented below (see Section 3.2) and due to the lack of dust correction should be treated as lower limits. In Figure 2.10 we present the difference of the metallicity of the gas and the galaxies as a function of the impact parameter. We find that in all cases the galaxy metallicity is higher than the gas metallicity lower limits. However, our detection threshold of emission metallicities goes down to 12+log(O/H) = 7.5, which corresponds to the absorbing gas metallicities.

Measuring the metallicity in two different spatial positions (in the associated galaxy and in the halo) can be interpreted as measuring a metallicity gradient. Resolved studies of nearby galaxies (like CALIFA or MaNGA) report negative ISM metallicity gradients in nearby star-forming galaxies (Sánchez et al., 2014; Ho, 2016; Belfiore et al., 2017). Similar studies of high-redshift lensed galaxies have shown a diversity of gradients in star-forming galaxies (Queyrel et al., 2012; Jones et al., 2013, 2015; Leethochawalit et al., 2016). Positive metal gradients can point to strong outflows expelling metals to the CGM or suggest inflows of a large portion of metal-poor gas.

Expanding the metallicity gradients from the ISM to larger scales likely probes the physics of the CGM. Gradients between absorbing gas and galaxies were so far measured for systems with impact parameter below 100 kpc (Péroux et al., 2012; Christensen et al., 2014) and have an average value of -0.022 dex kpc^{-1}. Combining results from absorption spectroscopy of gravitationally lensed quasars with measurements for star-forming galaxies, isolated spirals, quasar-galaxy pairs, and gravitationally lensed galaxies, Kulkarni et al. (2019) also reported a tentative anti-correlation between the metallicity gradient and the

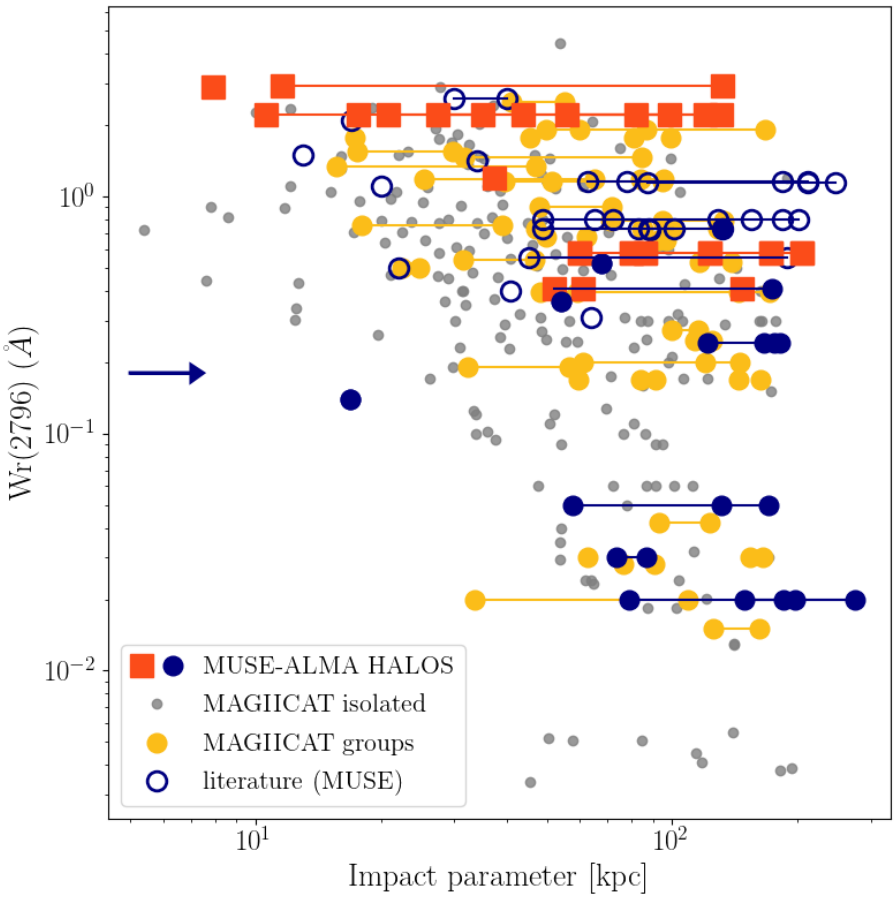

Figure 2.9: MgII λ2796 absorption line equivalent width versus the impact parameter. The filled blue circles are the galaxies identified in MUSE-ALMA Halos (groups linked with a line, absorbers selected based on their HI feature are marked with orange squares), empty blue circles mark systems from the literature (Schroetter et al., 2016; Bielby et al., 2017; Klitsch et al., 2018; Zabl et al., 2019), groups are linked with lines. The blue, horizontal arrow marks the limits of the Q1211z105 with no detected associated galaxy. The underlying points come from the MAGIICAT catalogues of isolated galaxies and groups catalogues in grey and yellow, respectively. Groups of associated galaxies alongside isolated systems define a region in parameter space below which no absorbers are detected.

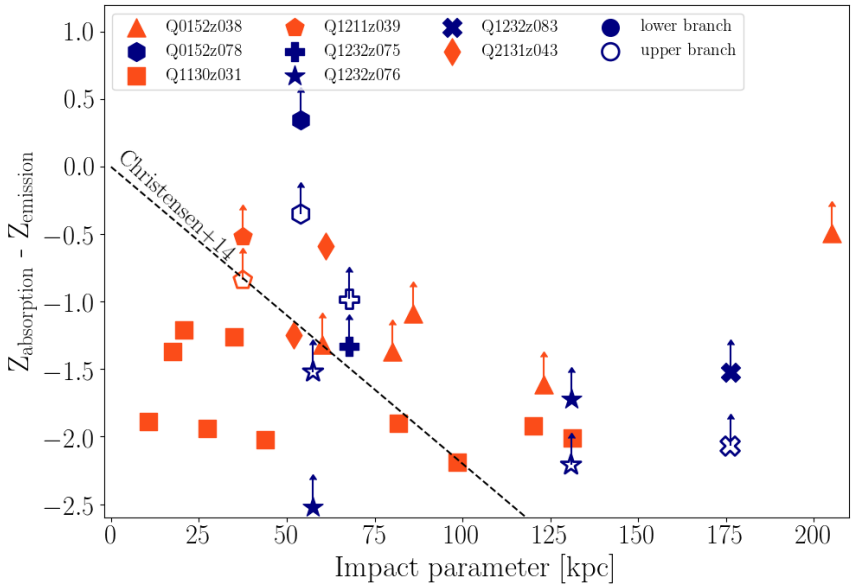

Figure 2.10: Difference in metallicity of the QSO absorption line and the metallicity from emission lines of the galaxies associated in redshift. Different symbols distinguish groups of galaxies associated with various absorbers,absorbers selected based on their HI feature are marked in orange. Filled symbols mark the lower branch metallicities or unequivocal values from the literature, hollow symbols mark the upper metallicity branch. The dashed black line marks the CGM gradient measured by Christensen et al. (2014).

metallicity at the galaxy centre: objects with low metallicity at the centre have positive metallicity gradients going outward.

We looked at the difference between the absorber and galaxy metallicity as a function of impact parameter (Figure 2.10). Note that our absorption metallicity measurements do not have detailed ionization correction, and additionally, we compare the values based on different spices: Fe for absorption and O for emission. We find no dependence of the metallicity difference on impact parameter and a larger scatter than Christensen et al. (2014) at redshift $0.1 < z < 3.2$. This is explained in part by the multiple galaxies associated with each absorber.

Hydrodynamical simulations like FIRE or IllustrisTNG (Muratov et al., 2015, 2017; Nelson et al., 2019), predict metals to be a tracer of the CGM structure following the flows of gas in and out of the galaxy. Metal-rich outflows are expected to originate in the centres of galaxies, introducing the metallicity differences with respect to the azimuthal angle between the quasar line of sight and the projected major axis of the galaxy. Future work will look these spacial variations of metallicity in more detail (see Péroux et al., 2016). To relate the gas probed in absorption with the galaxies in the group, a detailed component-by-component analysis provides a promising avenue (Rahmani et al. in prep., Rahmani et al., 2018a).

2.5 Conclusions

We report the first results of MUSE-ALMA Halos, a unique multi-wavelength data set aimed at understanding the complexity of the Circumgalactic Medium, the interactions with the galaxies and the environmental effect on the high HI column density absorbers. The survey consists of five quasar fields for which we identified 14 absorbers at redshift $z < 1.4$. We searched for the galaxies associated with the absorber within the MUSE field of view (covering 200-250 kpc in impact parameter, depending of the system's redshift) and within $\Delta v = 1000$ km s^{-1} from the absorber redshift. Each absorber has an HI column density measured either from the literature or from archival FOS spectra. We used MUSE IFU observations to identify galaxies associated with the gas seen in absorption, finding altogether 43 such systems yielding a 89 per cent detection rate among the non-targeted absorbers. We find that most absorbers are associated with multiple galaxies or groups, tracing overdensities of at least a factor of 10 relative to the field galaxies. The detection of multiple galaxies is directly related to the ability to map a large field centered on the QSO, which is a unique capability of MUSE covering impact parameters up to 250 kpc. In addition, the use of spectral PSF subtraction for low impact parameter cases and the sensitivity of MUSE resulting in the detection of low SFR object (SFR $>$ 0.01–0.1 M$_\odot$yr^{-1}) allow for much more complete study of the galaxy environment linked to the absorber. The observed absorbing gas may then probe not only a single CGM halo but most likely the intra-group material affected by intergalactic interactions. By studying the relative velocity differences between the gas and the associated galaxies, we find that the traditional way of assigning the absorber to the lowest impact parameter galaxy should be challenged. We find large dispersion in velocities between galaxies and absorbing gas, sometimes pointing to a more distant galaxy as the main candidate for the host. We compare the parameters of the systems in our sample with canonical relations of the absorbers and their host galaxies, N(HI)–b and $W_r(2796)$–b. We find that the $W_r(2796)$–b relation follows the general trend while defining an envelope of parameters below which no further objects are detected. Thanks to our multi-wavelength dataset we can compare the lower limits of metallicities of the absorbing gas with the metallicity of the galaxies as measured in the emission. We find that the overall metallicity is always higher in the associated galaxies than in the absorbing gas probed by the QSO. However, we are not sensitive to the low metallicity systems below 12+(O/H) = 7.5, and gas metallicities are lower limits. Given the complexity of the environment associated with these absorbers, we advocate the use of the kinematics information that could assist in associating individual absorbing component with each galaxy member. Comparison to the resolved kinematic and ISM metallicity of the galaxies could improve our interpretation of the absorption lines. With this study, MUSE-ALMA Halos, we show that combining absorption and emission studies of CGM and associated galaxies is a powerful tool to study the evolution of gas inside and outside galaxies. In forthcoming work we will study the molecular gas content of galaxies associated with the absorbers in our sample, using completed ALMA observations of all the quasar fields described in this work.

Chapter 3

ALMACAL project - surveying the Universe with ALMA

3.1 Introduction

Sub-millimetre astronomy targets the dark, cold Universe of molecular and atomic gas and dust emission at temperatures typically in the range of 10–100 K, typical conditions under which star formation processes proceed. Star formation can be traced on different scales, from local molecular clouds and protoplanetary disks to extragalactic galaxy-wide star formation, up to the epoch of reionization ($z \sim 8$).

The sub-millimetre is the high-frequency end of radio astronomy (100 – 1000 GHz) and thanks to interferometry, offers a detailed view of star formation processes in the local Universe as well as in the highest redshift galaxies.

In this chapter, I describe the ALMACAL project, the survey aiming at utilising the calibrator data for surveying the sub-millimetre sky with ALMA. I describe the properties of interferometric observations and the technical challenges and data processing of the ALMACAL survey, which is the backbone of the untargeted CO emission line survey described in Chapter 4. Additionally, the pipeline described in this chapter provides a basis for the extension of the pilot emission line survey to a full ALMACAL untargeted emission-line survey (see more in the Future Work in Chapter 5).

3.1.1 Sub-millimeter interferometric observations

Sub-millimetre astronomy uses similar antennas and observation techniques as radio astronomy. These high-frequency bands lay on the edge of the atmospheric radio transmission gap, with higher frequencies becoming more difficult to observe due to the increasing atmospheric opacity (Figure 3.2). The main component of the atmosphere blocking these frequencies is water vapor, therefore the sub-millimetre observatories are preferably located on high dry sites like the Plateau de Bure in the French Alps (NOEMA) or Chajnantor plateau in Chilean Andes (ALMA, Figure 3.1).

Figure 3.1: ALMA observatory, internationally run, sub-millimetre interferometer located on the Chajnantor plateau in the Atacama desert, Chile. *Image credit: ALMA (ESO/NAOJ/NRAO)/W. Garnier (ALMA)*

ALMA (Atacama Large Millimetre Array) is operated by the international partnership of European, North American and East Asian radio observatories under the Joint Alma Observatory (JAO). The interferometer is located on a high-altitude desert in the Chilean Andres (5000 m above the sea level), one of the best astronomical sites in the world. Its location is unique thanks to the very low levels of water vapor (below 1 mm Precipitable Water Vapor for 50 per cent of the time in the operational months between April and December). Water vapor is the most significant limiting factor in the quality of observations as the water molecules in the atmosphere easily absorb the sub-millimetre radiation, decreasing the signal as well as increasing the noise by the thermal radiation. ALMA observatory consists of the main array of 54, 12-m antennas, four additional 12-m antennas (the "total power" array) and additional twelve 7-m antennas forming the Atacama Compact Array (ACA). The antennas can be moved to the different locations on the plateau, adjusting the configuration and the maximal distance between the antennas, which can reach 16 km. Currently, ALMA operates in seven frequency bands (Table 3.1).

Interferometry, a widely used the technique in radio astronomy, provides high sensitivity and angular resolution crucial for the detailed study of the star formation processes in the

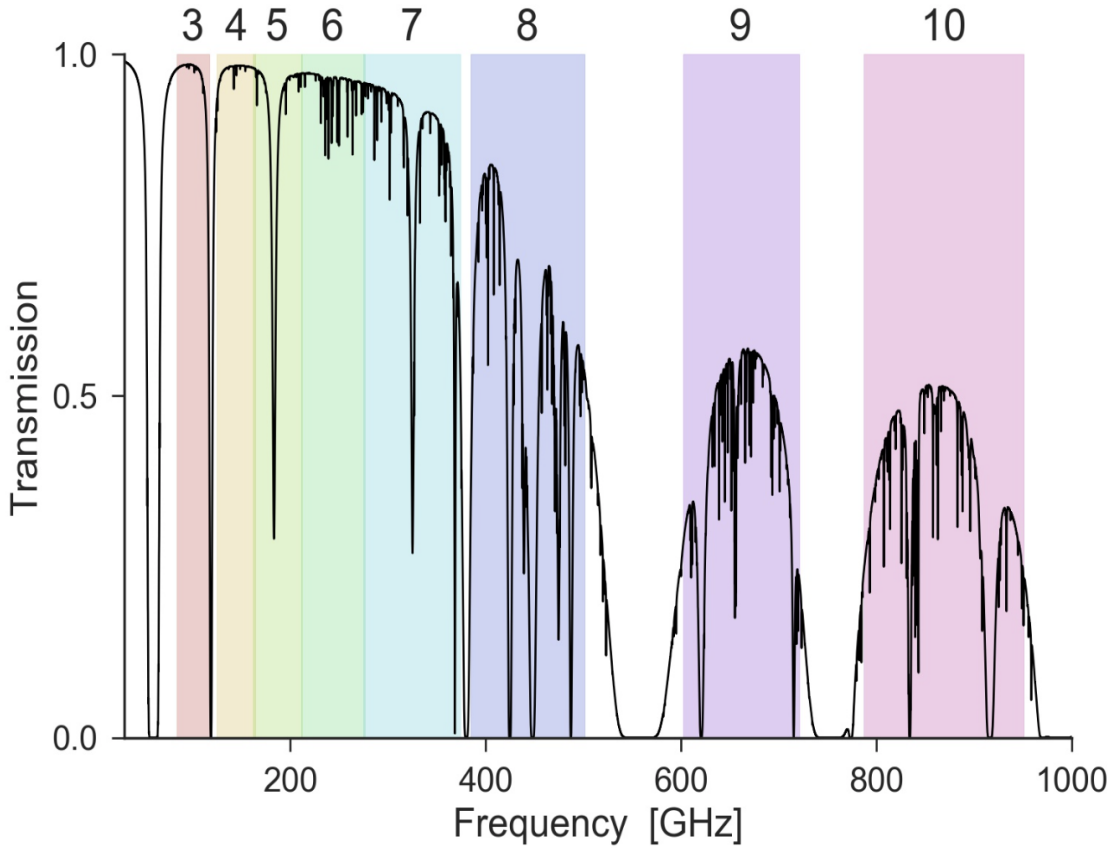

Figure 3.2: The atmospheric transmission spectrum at sub-millimetre wavelengths. The high transmission zones define the frequency bands (from 3 to 10, see Table 3.1) available for observations with ALMA. *Image credit:* `science.nrao.edu`

Milky Way as well as in other distant galaxies. An array of small telescopes is connected into one big virtual telescope, providing an angular resolution equivalent to the observations of the single telescope of the diameter equal to the largest distance between the antennas. All antennas are working together, observing the same object simultaneously. The data from all antennas are accumulated and combined, accounting for the phase shift between the antenna pairs. The resulting direct observation is the interferometric pattern of the signal from all of the antennas.

The physical distance between two telescopes defines a singular baseline - 2D vector representing the distance between each antenna pair. Measured amplitude and phase of the interferometric signal of the pair define the value of a corresponding visibility point. The measurements from each of the hundreds of possible antenna pairs altogether create a *uv*-plane (Figure 3.3). The *uv*-plane is a Fourier plane of spatial frequencies, representing the distance between antenna pairs, with the largest baseline defining the angular resolution of the image (the summary of the basic terms used in the interferometry is presented in

ALMA band	frequency [GHz]	primary beam [arcsec]
Band 3	84 - 116	62
Band 4	125 - 163	44
Band 5	163 - 211	34
Band 6	211 - 275	27
Band 7	275 - 373	18
Band 8	385 - 500	14
Band 9	602 - 720	9
Band 10	787 - 950	7

Table 3.1: Frequency coverage and primary beam angular size in each of the ALMA bands. Columns: (1) ALMA band, (2) frequency coverage of the ALMA band, (3) The size of primary beam (corresponding to the field of view) in arcseconds. *Source:* `almascience.eso.org`

Basic terms used in the interferometry	
baseline	The physical distance vector between any two antennas in the array.
uv **plane**	The Fourier plane of the observation representing the distance between all antenna pairs in 2D.
visibility point	One Fourier component of the source, corresponding to one baseline. Visibility is the measurement of the signal received by these two antennas.
array configuration	Antenna layout, can be more compact or extended, defined by the largest baseline. Defines the spatial resolution.
primary beam	Spatial resolution of one antenna, defines the field of view. Sensitivity is not uniform over the f.o.v. and falls of following the Bessel function.
synthesized beam	Spatial resolution of the array. The analogue of the PSF (seeing) in optical astronomy

Table 3.2: Interferometry definitions

Table 3.2).

Filling the *uv*-plane with visibilities can be achieved either by increasing the number of antennas or through the observations of the target at different times of the day, changing the orientation of the baselines toward the observed source, resulting in the arc-like structures of visibilities on the *uv*-plane (Figure 3.3). Eventually, a set of visibilities is the result of the interferometric observation. The analysis of the astronomical signal is already possible at that stage, however, it is more straightforward to analyze the reconstructed astronomical image. The Fourier Transform of the *uv*-plane leads to the reconstruction of the image from the combined visibilities. The total flux density of the image comes from the Fourier Transform of the 0th baseline (all antennas in one place), in practice impossible to achieve due to the physical size of the antennas. The total flux observation is supplemented by pairing with the single dish (NOEMA) or a compact array (ALMA).

The field of view of a radio or sub-millimetre telescope is limited by the primary beam, the 2D response of the radio antenna to the incoming signal. The primary beam size dependents on the diameter of the individual antennas in the array (D) and the observed wavelength λ:

$$f.o.v \sim \frac{\lambda}{D} \tag{3.1}$$

ALMA's field of view varies from 1 arcmin in Band 3 (lowest available frequencies, 84–116 GHz) to 10 arcsec for high frequencies in Band 9 (602–720 GHz). All frequency coverages and probed fields of view for different ALMA bands are summarized in Table 3.1. The resolution element or the point-source response to the aperture synthesis of the combined array is the **synthesized beam**. It is the radio analogue of the Point Spread Function in optical astronomy and its size is proportional to:

$$\sim \frac{\lambda}{B}, \tag{3.2}$$

where B is the maximal baseline of the array configuration.

The important part of the interferometric observations, which could take up to 20–30 per cent of the observing time, are the calibrations. Especially in interferometric observations, repeated calibration observations are required to ensure the right combining and scaling of the signals of the antenna. Ideal calibrator sources are bright sub-millimetre point-like objects that do not show any large scale structures on the probed spectral scales. Additionally, an ideal calibrator should have a constant flux, or at least vary only slowly in brightness. The set of calibration observations are especially important for the ALMACAL project and the study described further in this and the next chapter (see Section 3.2). The main calibrations of radio interferometric data include:

- **bandpass calibration** - phase and amplitude measurements as a function of frequency; takes out spectral response from the data,
- **phase calibration** - measurement of the phase changes with time of the observations,

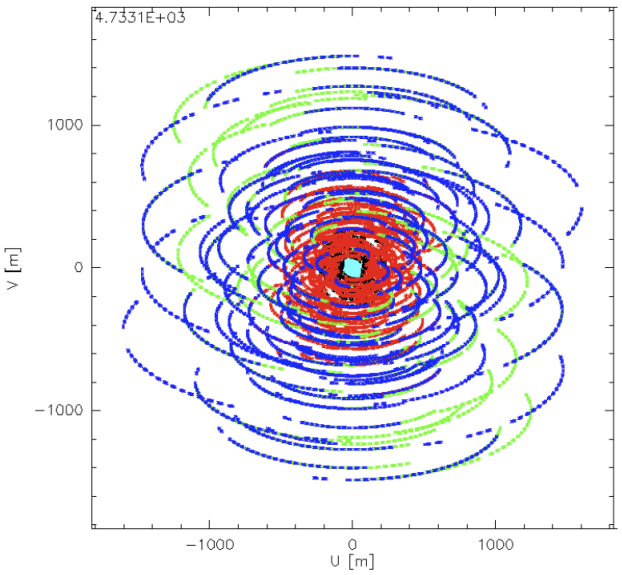

Figure 3.3: An example of the well-sampled observations in the *uv*-plane. The arcs are the effect of the Earth's rotation changing the angle at which each pair of antennas see the object. By repeating observations on different times of the day, the *uv*-plane can be filled with more visibility points. *Image credit:* `science.nrao.edu`

- **amplitude calibrations** - measurement of change of the signal amplitude with time,
- **flux calibrations** - calibrations to absolute flux. Requires well-known, non-variable objects of flat spectrum.

Additionally, the state of the atmosphere is constantly monitored to correct the measured phase also for atmospheric scintillation and water vapor variation.

Finally the calibrated measurements can be converted into 3D datacube containing spectral and spatial information of the observed source.

ALMACAL

Figure 3.4: The logo of the ALMACAL project. The summary of the project including a publication record, is available at `almacal.wordpress.com`.

3.1.2 Motivations of the ALMACAL project

ALMA is a powerful tool for the exploration of the cold distant molecular Universe. Although ALMA is very sensitive, due to its small field of view, the interferometer is not ideally suited for large sky surveys. Studies of targeted objects have shed new light at the molecular gas in galaxies and the nature of the ISM in systems on different redshifts. ALMA has pushed the observations of galaxies in the early Universe to a record-breaking $z = 9.1$, (Hashimoto et al., 2018). However, all these achievements are characterised by targeted surveys and require a large time allocation to be completed.

The time available through ALMA Large Programs seen against the amount of integration time required to trace fainter objects over a large region of the sky pose a challenge to use ALMA as a survey machine. Interestingly, an interferometer requires recurrent acquisitions for the calibration sources (summary in Section 3.1.1). These corresponding calibration data are always taken alongside science observation and keep accumulating with each observing cycle.

The ALMACAL project was established to utilise this vast dataset and address astronomical questions that require a more survey-like approach: the sub-millimetre dusty galaxies number counts, the nature of bright sub-mm quasars (which are the majority of ALMA calibrators) or an untargeted search for CO emission lines (the full list of the ALMACAL science cases is described in Section 3.2.4). With the use of the calibrator fields, ALMA can, in effect, be transformed into the survey instrument without additional costly time allocation for the observatory.

3.1.3 Motivation of ALMACAL-CO project

ALMA is the ideal instrument for studies of CO emission from galaxies across cosmic time. Figure 3.5 shows the coverage of the lowest CO transitions (Table 3.3) in different ALMA bands. CO emission lines trace the molecular gas, linked directly to the star formation processes (see Introduction Section 1.2.2).

CO emission line studies so far were conducted using two approaches: through targeted surveys of samples of star-forming galaxies, assumed to be representative of molecular gas reservoirs in the Universe (Tacconi et al., 2020), or in an untargeted way, searching for any emission lines in the area surveyed (Decarli et al., 2016, 2019; Riechers et al., 2019). The latter was so far conducted over small cosmological fields, that have substantial multi-wavelength coverage but, as they typically cover a small consecutive patch of the sky, can be sensitive to the effects of cosmic variance. To supplement the CO ALMA observations conducted so far we chose to use the ALMACAL data to design a CO untargeted emission line survey. Thanks to the uniqueness of the ALMACAL dataset we can study galaxies over a wide area in the directions widely separated over the southern sky and are less sensitive to the effects of cosmic variance. Additionally, the ever-growing database of new observation provides an opportunity to update the search and to continuously refine the constraints on the molecular gas evolution.

In this chapter, I describe the technical challenges behind the archival data-mining in

the ALMACAL-CO survey. In Chapter 4 we present the preliminary results of the pilot survey demonstrating the new approach.

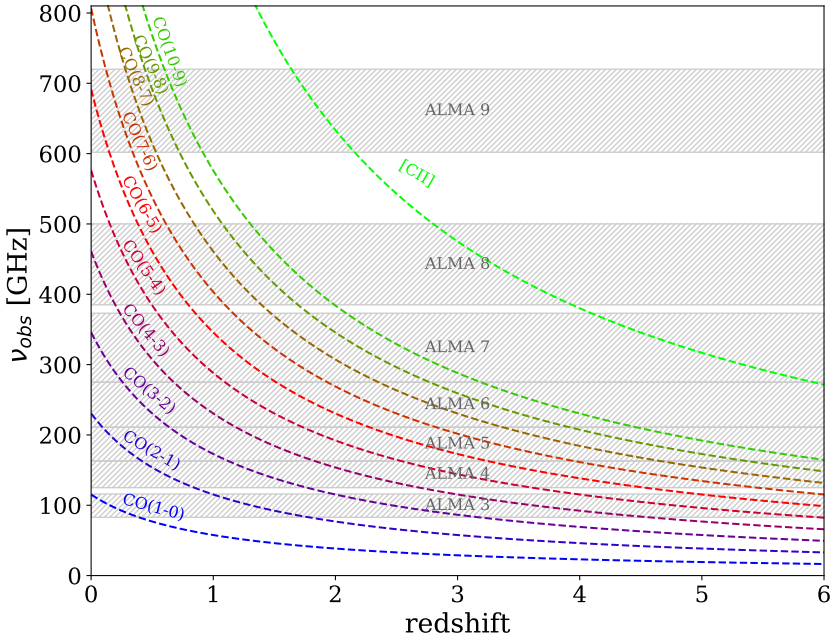

Figure 3.5: The coverage of the CO transition line ladders in the ALMA frequency Bands with redshift. The frequency coverage makes observations available of low–J lines in low redshifts and following the evolution of molecular content of galaxies in higher redshifts for higher–J transitions. Additionally for redshift about $z > 3$ the bright [CII] line becomes observable providing an additional bright signature for star-forming galaxies in the early Universe.

3.2 ALMACAL: Using ALMA calibrator for scientific purposes

The ALMACAL is a project utilising the vast archives of ALMA calibrator data. Due to the nature of the interferometric observations, each science observations is accompanied by numerous calibration observations, as described in Section 3.1.1. From the beginning of science operation of the ALMA observatory, the accumulated calibration data amount to over 4000 hours of observations, the equivalent of half ALMA's yearly observation cycle. This amount of data exceeds any allocated Large Science program by a factor of 10–100.

In each self-contained block of ALMA science observation (called the execution block, EB) about 20–30 per cent of the time is spent on calibrators. As described in Section 3.2, the calibrations include measurement of the flux calibrator, bandpass response calibrator and the amplitude and phase calibrations for the scientific source.

CO J_{up}	Rest-frame ν [GHz]
1	115.27
2	230.54
3	354.90
4	461.04
5	576.27
6	691.47
7	806.65
8	921.80
9	1036.91
10	1151.99
[CII]	1900.5

Table 3.3: Rest frequencies of CO transitions. Columns: (1) The CO transition J_{up} number, (2) Rest frame frequency of the transition in GHz.

In this section, I describe the properties of ALMACAL dataset - the parameters of the data as well as the characterisation of the ALMA calibrators.

3.2.1 ALMA calibrators

The ALMA calibrators are sub-millimetre bright point sources (flux density typically in the range of 0.1–1 Jy), that mostly come from the sub-millimetre all-sky Australia Telescope 20 GHz Survey (Ekers et al., 2007). These bright sources are mostly extragalactic, however several solar system objects (Saturn, Uranus, Jupiter etc.) are also used as flux calibrators.

In ALMACAL we focus only on the extragalactic calibrators. Over 880 sources are distributed over the ALMA sky (Figure 3.7) and cover about 800 arcmin^2 on the sky in Band 3 (down to 1 arcmin^2 in Band 9, see Figure 3.6). Since the beginning of ALMA science operation, the number of calibrators has gradually increased, but at present, new calibrator fields are introduced at a slower pace.

97 per cent of the calibrators are characterised by a flat-spectrum consistent with different types of blazars, a type of AGN which mostly featureless optical spectra makes it difficult to obtain the redshifts. Archival redshift estimates are available for 78 per cent of the sample, however, only a fraction of them is confirmed spectroscopically. Most of the calibrator have redshift between $0.5 < z < 1.0$ with a small percentage expanding towards $z \sim 3.5$. Precise redshift measurements of the calibrators are especially important for the absorption line surveys in their spectra. An ongoing effort is conducted to complete the

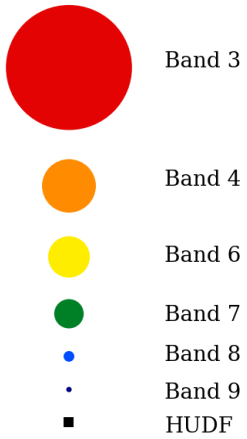

Figure 3.6: Illustration of the summed sky coverage of ALMACAL data in different ALMA bands. The covered sky area ranges from about 800 arcmin^2 in Band 3 to 1 arcmin^2 in Band 9. The angular size of the Hubble Ultra-deep Field (HUDF, black square at the bottom) is shown for comparison.

spectroscopical redshift measurements through optical observation.

Bonato et al. (2018) looked at the AGN properties of the calibrators. The nature of calibrator observations, repeating the measurements every several days in different frequency setups, allows for the time domain light-curve sampling as well as for measuring the SEDs of the objects. Typically for AGNs, the blazars show strong flux variations with time also in the sub-millimetre regime.

Which calibrators are chosen for each science observation depends on their position in the sky. They should be relatively close to the source, therefore calibrators in some regions of the sky are covered more often than others, depending on their proximity to interesting science objects or deep cosmological fields.

The sub-millimetre brightness of blazars is the result of the geometrical orientation of their jet towards the observer. They are not particularly massive, unique objects in the Universe around which galaxies would significantly cluster (unlike the massive radio galaxies). Using the blazars in an untargeted survey is therefore not introducing the bias of source clustering.

3.2.2 ALMACAL data reduction and database

Each self-contained execution block, of typically $\sim$ 45 min, executed with ALMA is accompanied by several calibration observations of typically 2–3 calibrators. Depending on the

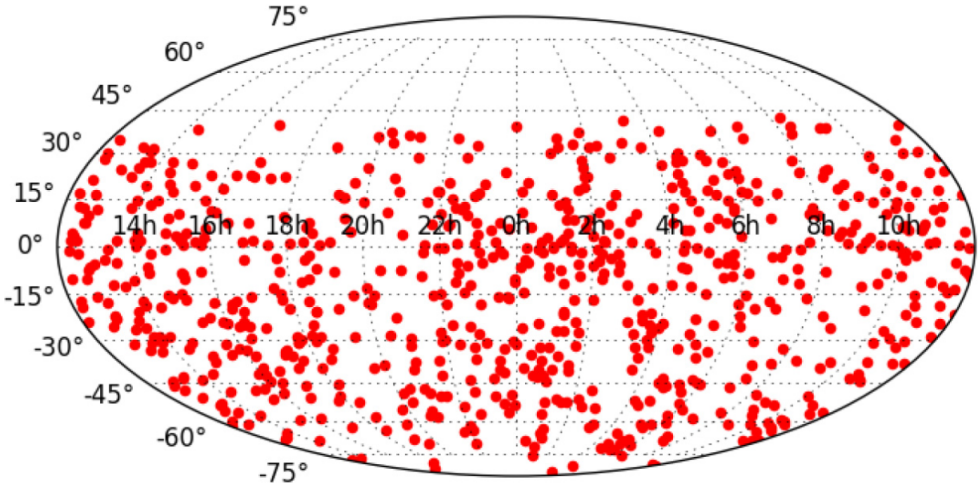

Figure 3.7: Distribution of ALMACAL fields (ALMA calibrators) on the sky (figure from Bonato et al., 2018). The size of the points does not represent the physical scale. The random distribution of fields over the sky makes the ALMACAL surveys less prone to the effects of cosmic variance.

duration of the project these calibrations can be re-observed in the same setup multiple times during the PI science observations executions. Additionally, the same calibrator can be re-observed for other science purposes increasing the total observation time.

According to the ALMA operations plan, all calibrator observations become immediately publicly available. In practice, these data are embedded in the PI data package and requires download of the full science package. In the ALMACAL survey, we take advantage of the help of the ESO ALMA ARC in retrieving the calibration data and generate our database consisting exclusively of these observations.

To start with, data packages are extracted from the ALMA archive together with the suitable PI pipeline script for the reduction and calibration. The data are reduced with the script for PI, the science data is removed afterwards. Science observations are not used in the ALMACAL project.

Additional calibrations must be applied to the bandpass calibrator, which is typically not flux calibrated. The data are self-calibrated on the central calibrator source. The self-calibration provides the best possible calibration results and thanks to the presence of bright point source in the centre of the field of view, can be used on all our dataset. It is carried out in two steps: in phase-only and then phase with amplitude calibration.

To use the ALMACAL data for a deep survey, the bright point source (calibrator) must be removed from the dataset. As the calibrators are chosen to be point sources without complicated internal structure, and preferably without any spectral features, the removal of the point source is relatively straightforward and efficient. Finally, all data are set to

the same spectral resolution of 15.625 MHz. After these steps, the data are ready to be imaged and analysed.

3.2.3 ALMACAL properties

ALMACAL data parameters always depend on the requests that PIs of the corresponding science projects. As calibrators are observed recurrently in the same setup as the science data, we have no control, especially over the frequency coverage. Data are also taken in different array configuration with different spatial resolution, sometimes providing high-resolution images. The spatial resolution of the data can be chosen during a data processing by choosing which antennas on the uv-plane to be included in the final dataset.

The ALMACAL database includes calibration data starting form ALMA Cycle 1 and is regularly updated with the new deliveries of calibrator observations. However, in Cycle 1, not all antennas from the configuration were yet available and that data is less valuable than the one from the following cycles. Currently, in 2020 we are in Cycle 7, ALMACAL accumulated already seven years of data.

The majority of ALMACAL datasets are in Bands 3, 6, 7, where the observations are undertaken most often. Observations in high-frequency bands (bands 8, 9 and 10) are more difficult to conduct due to increasing atmosphere opacity to the high frequencies. In results, high-frequency bands are observed less frequently than ones covering lower frequencies.

The total number of observing hours is over 4500 h from all calibration fields. Most of the calibrators are observed in Bands 3,6,7 and that also is reflected in the total integration time in these bands, as it is the largest. Integration times of the calibrations in similar setup differ from 10 minutes to a few hours. The summary of the number of calibrators and integration times is shown in Table 3.4 and Figure 3.8.

3.2.4 Examples of the ALMACAL science results

The specifics of the ALMACAL data, a sky distribution and repetitive observations of the same fields allow for conducting surveys of the sub-millimetre sky. Since the start of the ALMACAL project in 2016, a number of science cases for the project have been proposed. The original idea was the use of the multiple observations of calibrators to obtain ultra-deep continuum images to search for dusty star-forming galaxies at all redshifts.

The first search for the dusty bright sub-millimetre galaxies (SMGs) was done by Oteo et al. (2016, 2017), who looked at the 69 calibrator fields with deep images in ALMA Band 6 ($\sim$ 1.2 mm) and band 7 (870 μm) available at that time, providing first constraints for faint systems in 870 μm regime. The extended studies of the number counts of SMGs were continued with an updated dataset in Band 8 Klitsch et al. (2020). These findings are consistent with resolving 100 per cent of the cosmic far-infrared background with these dusty objects at 650 μm.

The calibrators (most of which are blazars) themselves are interesting sources for studying AGN physics in the sub-millimetre. Husemann et al. (2019) studied the CO emission

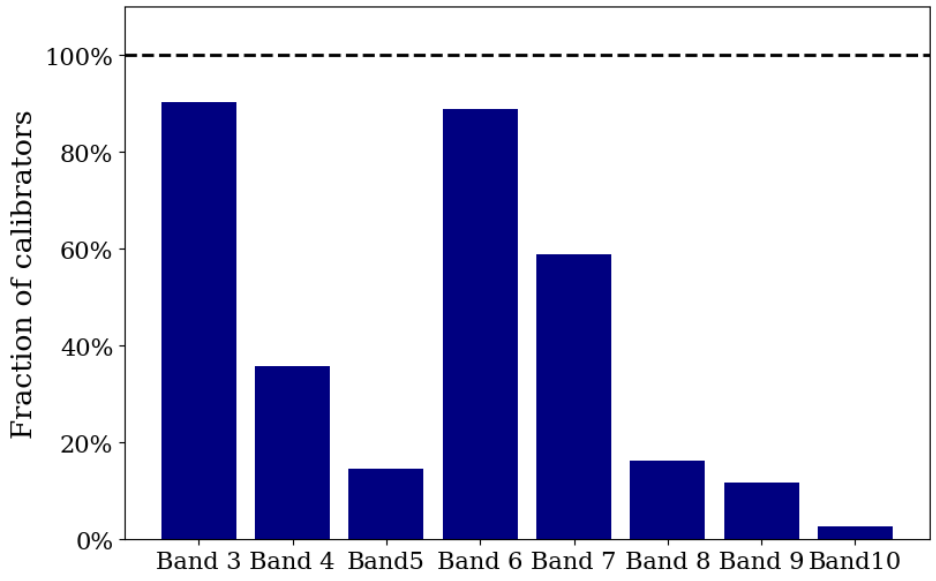

Figure 3.8: The fraction of ALMACAL calibrators covered in different ALMA frequency bands. Most observations are conducted in Bands 3, 6 and 7, where 60–90 per cent of calibrators are covered. Fewer observations are available for high frequencies, in bands 8, 9, 10 which are mostly affected by atmospheric conditions.

from the host of the popular radio source 3C273, which is also an ALMA calibrator, detecting the faint CO emission coming from the galactic disc. The finding illustrates the large dynamical range that can be reached with the ALMACAL data. Bonato et al. (2018) looked at the calibrator data in a time domain, creating the light-curves of all blazars. ALMACAL observations, repeated over the years allow for multi-year follow up of the AGN variability, poorly studied aspect of AGN science.

Blazars can be also used as distant point sources in search of intervening absorption lines. Most of the quasar absorption systems were discovered in UV spectroscopy of intervening absorption lines of hydrogen and metals. Hunting for the absorber's host galaxies at high redshift in the optical is challenging due to their low brightness and the fact that no significant bright lines other than Lyα are available in their spectra (see Chapter 2). Instead, the search for bright sub-millimetre lines offers the detection of the absorbers host in CO or [CII] and help quantify the molecular gas content of such selected objects (Neeleman et al., 2018). In ALMACAL, a cross-correlation of known absorbers in the calibrator spectra with the available frequency coverage resulted in the CO emission line detections from several absorbers host galaxies (Klitsch et al., 2018, 2019a). Thanks to the extended frequency coverage in some fields it was also possible to observe multiple CO lines and determine the CO line ratios, challenging to study the Spectral Line Energy Distributions (SLEDs). Our studies have shown a higher ISM excitation levels for the hosts of these absorbers than in the Milky Way (Klitsch et al., 2019a).

The ALMACAL database allows for studying the evolution of global quantities like the

ALMA Band	n_{cal}	T_{tot} [h]
Band 3	794	787
Band 4	315	1664
Band 5	129	64
Band 6	783	1100
Band 7	518	783
Band 8	143	120
Band 9	103	80
Band 10	24	8
Total		4606

Table 3.4: The summary of the calibration data included in ALMACAL, covering only extragalactic calibrators. Table columns: (1) The ALMA band, (2) Number of calibrators covered in that band, (3) The total integration time of the calibrators covered in the band in hours. Calibrators differ in the frequency coverage and are not always observed in all bands (hence $n_{cal} < 880$).

molecular gas mass function. The measurements of H_2, the most abundant molecule in the Universe and main molecular mass constituent, cannot be achieved directly in extragalactic sources due to the low brightness of its vibrational transitions. Instead, the typical way of tracing the molecular gas content of galaxies across redshifts is through bright CO lines. ALMA observations can cover a range of CO lines at different redshifts as shown in Figure 3.5. This allows us for probing the molecular gas content across redshifts. CO emission line search (ALMACAL-CO survey) is the topic of this and the next chapter, while the complementary CO absorption survey towards the ALMACAL calibrators is described in Klitsch et al. (2019b).

3.3 The ALMACAL-CO: automatised emission-line search in the ALMACAL database

By studying the bight CO emission in galaxies we quantify their molecular gas content, in turn deriving the molecular gas density of the Universe. An unbiased way to study the molecular content of galaxies across cosmic time is through untargeted emission-line surveys (ASPECS - Decarli et al. (2019), COLDz - Riechers et al. (2019)). To date, such surveys have been limited to single fields, covering a small part of the sky, being prone to the effects of cosmic variance. To quantify the molecular gas budget of the Universe and the influence of the molecular gas content of galaxies on the star formation history of the

Universe (SFH) we need a statistically significant sample of CO-bright galaxies, unaffected by the effects of cosmic variance.

To complement the deep pencil beam surveys I conducted an archival data survey, utilizing the ALMA calibration data in the ALMACAL project thereby constructing a wide-area untargeted CO emission line survey, the ALMACAL-CO. In the ALMACAL-CO pilot, I chose the fields with the longest total integration times (38 fields, observation over 40 minutes) and detected eleven emission lines (Hamanowicz et al. in prep). Since auxiliary data are not available for the ALMACAL-CO fields, we could not easily define the redshifts and uniquely identify the CO transitions. Based on the population of galaxies from the Shark semi-analytical models (SAMS, Lagos et al., 2018), I assigned a CO transition probability to each detection.

In this section, I describe the details of the technical background behind the selection and analysis of the data. The ALMACAL-CO pilot project was the opportunity to develop the data mining methods and reduction and analysis pipeline of the ALMACAL data which automatise the search for emission lines in the data. The automatization and the pipeline chain from the database to the final list of sources allows for the quick evaluation of the data and is a preparation for the next step after the pilot project, the ALMACAL-wide search for emission. The detailed description of the projects and the result of the pilot study are presented in Chapter 4.

3.3.1 Selection of the deep sample

In this subsection, I describe the selection procedure applied to create the ALMACAL-CO pilot dataset. The summary of the whole data processing from the data selection to the final list of candidates is presented in Figure 3.14.

To choose the best subset of the ALMACAL-CO project we mined the ALMACAL archive, in search for suitably deep observations. We required our data to be homogeneous, as any noise variation or concatenation of the mismatched data could increase the r.m.s. of the observations making the detection of faint CO emission lines more challenging.

Singular calibrator observations are of the order of several minutes but are repeated multiple times during the execution of the science project, resulting in long observations over particular calibrator fields. I searched the ALMACAL database for such cases to obtain the deepest possible observations in the same frequency coverage. I excluded the datasets that were taken in spectral scan mode. For the spectral windows (spw) I chose only the ones broader than 0.5 GHz. Calibration observations are also performed on the ACA array, consisting of 7-m antennas, however, these data are not a part of the ALMACAL.

To select the deepest observations, I inspected the distribution of the total integration time of observations over the same frequency, choosing the datasets of at least 40 minutes of exposure time.

Before combining the data to create datacubes I imaged all individual datasets separately and inspected them by eye. Datasets with visible artefacts from non-perfect continuum subtraction of calibration imperfections were removed. I classified images in the three categories: clean data, quasar residuals (bright source in the centre of the image) and

artefact (usually geometrical patterns). Examples of these three types of files are shown in Figure 3.9. All datasets with artefacts were removed from the sample and the ones with quasar residuals were flagged for further masking in the next reduction steps.

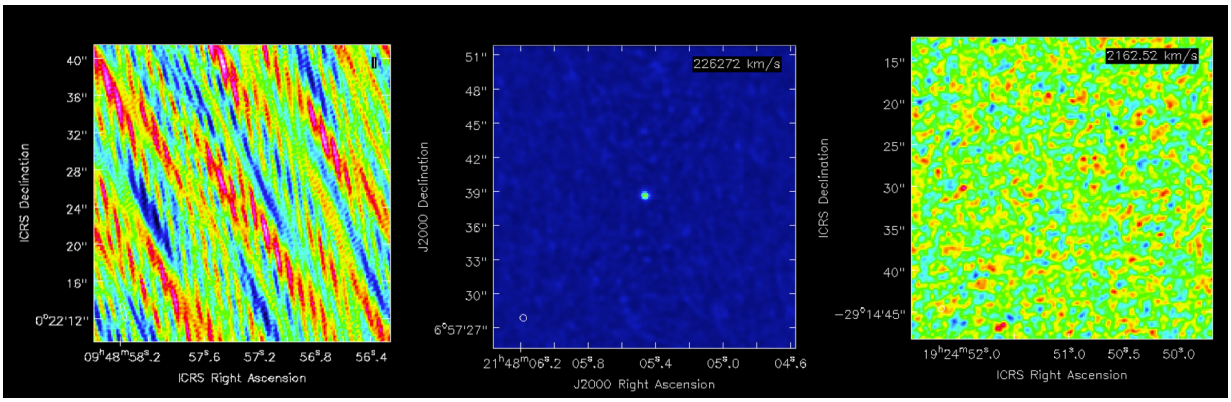

Figure 3.9: Example of the images made from the single datasets from ALMACAL-CO sample. Left: the observations classified as an artefact, such data was removed from the sample. Middle: The visible residuals of the central calibrator, due to the imperfections in point source removal. In case of such data, the central region equal to two resolution element was masked. Right: the noise dominated, typical ALMACAL-CO dataset.

3.3.2 Creation of datacubes

All of the data processing of the ALMA data files is done with the CASA software (Common Astronomy Software Application) used for the reductions of the radio interferometric data.

Continuum correction

The ALMA calibration data is dominated by the central bright source, the calibrator. Although the calibrator point source is removed in the process of the ALMACAL data reduction, the residuals of imperfect subtraction can influence the images. I corrected for this by using the *uvcontsub* routine, which fits a first-order polynomial in the frequency space. I applied this correction to all dataset, as a variation of the continuum could affect the search for faint emission lines. To create the datacubes I first concatenated all datasets corresponding to the same field and frequency setup using the *concat* routine.

Choosing datacube properties

Creating data cubes for interferometric data requires the choice of the image size, corresponding to the part of the primary beam I include in imaging, and pixel size. The field of view of an interferometer is limited by the size of the primary beam, which depends on the observed frequency. However, sensitivity is not uniform over the field and falls off from the centre of the field, following a Bessel function (which can be approximated with

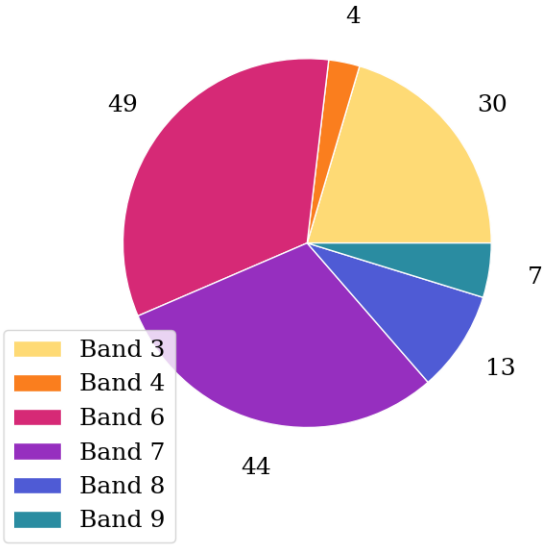

Figure 3.10: The number of ALMACAL-CO pilot cubes covered in different ALMA bands. The majority of cubes in our sample come from Bands 3, 6 and 7, reflecting band coverage of the total ALMACAL database.

a Gaussian). Typically to set the size of the image, the FWHM of the Gaussian beam is used. For 300 GHz and 12-m antennas that gives a 21 arcsec field-of-view.

Custom functions from CASA Analysis Utilities provided the best parameters of image and pixel size for a given dataset. To limit the storage size of the cubes and at the same time keep all information, I limited the total number of pixels to the optimal three pixels per beam.

Spatial resolution of the data

In the CO emission line search, I want all candidates to remain unresolved or poorly resolved to optimise the detection process. The CO emission in L_* galaxies in known to be concentrated in the central part of the galaxy (e.g. Villanueva et al., 2017; González-López et al., 2017). Assuming that the observed CO emission region is concentrated within $\sim$ 4 kpc, for redshifts $z = 0.5 - 2$ this will not exceed 0.5 arcsec, and hence that was the minimal spatial resolution limit I applied to the data.

The resolution of the data is defined by the synthesized beam, depending on the length of the baseline between the antennas with the largest separation. The data could come from different antenna configurations resulting in varying spatial resolutions. I placed a limit on the spatial resolution with the parameter *uvtaper*. This parameter controls the radial weighting of visibilities in the *uv*-plane through the multiplication of the visibilities by the Fourier Transform of an elliptical Gaussian, in practice down-weighting the largest baselines.

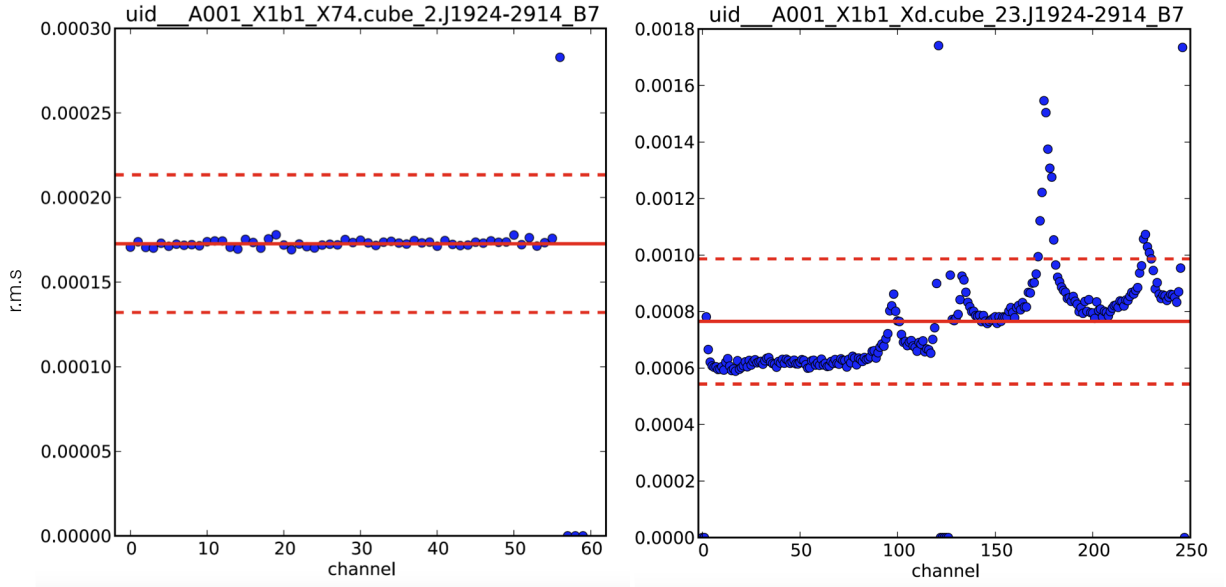

Figure 3.11: Examples of the r.m.s. per channel diagnostic plots for two data cubes from ALMACAL-CO pilot sample. The solid red line marks the median r.m.s. value across channels, the dotted red lines mark 1σ deviation from the median. On the left panel, the example shows a typical r.m.s. profile we considered good, on the right - the bad case of a noisy cube with irregular and significant variations of noise across channels. The channels which r.m.s. equals zero are the masked channels with the flux of zero. For source finding, we flagged all channels for which r.m.s. exceeding the 1σ deviation from the median noise value.

Datacube creation

Datacubes from the concatenated datasets were formed in CASA with the task *clean*. In constructing the datacubes I chose to use an equal frequency resolution of each cube ($\sim$ 15 MHz per channel), irrespective of the native resolution of the *uv*-data. Each visibility point was multiplied by a weighting factor, which determines which spatial scales are emphasized. I chose the Briggs weighting with the robustness parameter of 0.8. The higher the robustness parameter, the closer the weighting is to natural weighting, where all visibility points are treated in the same way.

3.3.3 Characterisation of the data

After creating all of the datacubes, we finalised the sample of ALMACAL-CO pilot to 147 cubes over 38 fields. The number of cubes, and also the frequency coverage, differs between the fields: from 23 cubes (in J1924–2914 field) down to single cube per field for five fields. We have cubes from all available ALMA bands in our sample (apart from very high-frequency Band 10 for which ALMACAL has fewer observations Figure 3.10). Each cube is based on at least 40 minutes of integration time. In the process, we had to

remove some cubes, due to strong r.m.s. variations, continuum subtraction artefacts, or other technical issues.

The total area of the sky covered by all the fields is 7.55 arcmin^2 and the probed volume is 216,470 Mpc^3. The detailed volume calculations of the survey are described in Section 3.4. 33 cubes had central source subtraction residual, which were masked by blanking the central part with a 5x5 pixel rectangular mask.

Variation of the r.m.s. between spectral channels influence the results of source detection. The cubes with huge variations of the r.m.s. profile per channel are likely to be difficult to process for source finding and additionally it will be difficult to quantify the reliability for sources coming from channels with highly variable noise properties. The examples of the r.m.s. variations in ALMACAL-CO cubes are presented in Figure 3.11.

We inspected the r.m.s. variations between channels and calculated the median value and standard deviation of the noise profile. These statistics were later used to list the outlier channels, for which the noise levels deviation from the median exceeded 1σ. All cubes with the r.m.s. value above 5 mJy/beam per channel was excluded from the sample. In the process, we found several cubes with exceptionally high r.m.s. values reaching 1 Jy/beam and removed them from the sample.

3.3.4 Source finder softwares

In this project, we tested two source finders originally developed for 21-cm HI radio surveys: Duchamp (Whiting, 2012) and SoFiA (Serra et al., 2015). Both were built for the large, sky-wide surveys for radio 21-cm HI emission but can be adapted for the use of the unresolved source searches with ALMA. Duchamp is no longer under development, but SoFiA, its successor, is constantly upgraded. The comparison of the properties of two source finders is summarized in Table 3.5.

Interferometric data cubes have certain characteristics that have to be taken into account for the effective source finding. Primary, for the search of unresolved sources, one must adopt the resolution element, the size of the synthesized beam, as the lowest possible spatial size of the source. Secondly, due to the noise fluctuations across the channels, narrow single-channel detections are unlikely to be real. At the same time, we expect sources to vary significantly in spectral width: from tens to hundreds of km s^{-1}. The source finder must be able to probe the spectrum at different spectral scales and be sensitive to this wide range of possible signal widths.

Most of our detections are expected to reside in the low signal-to-noise regime. The source finder tool should implement statistical criteria helping to disentangle the possible candidates from noise peaks. A convenient way of presenting the results is important for the verification of the suggested source candidates, both spatially and spectrally. We review the properties of the two tested source finders below, highlighting the criteria described above (see Table 3.5).

Duchamp

Duchamp is a three-dimensional source finder developed by Matthew Whiting at the Australia Telescope National Facility, CSIRO Astronomy Space Science (Whiting, 2012). Duchamp is designed to search for emission lines in frequency-based radio data cubes. In the process of source finding, no prior shape of the detection is assumed, allowing for different detection shapes. The source finder performs the three-dimensional search on the FITS data cube, producing a catalogue of sources detected above the signal-to-noise threshold, accompanied by flux maps and spectra of the detection. The user can decide alternatively to apply a spectral or spatial smoothing for the data. During the search, Duchamp checks each 1D spectrum form each pixel and then examines the pixels around the primary pixel detection to identify more spatial pixels belonging to the candidate.

The main feature of the finder, to tackle the issue of looking for a few channel detections in noisy data, is the wavelet reconstruction method. In this method, the subsequent set of wavelet functions are fitted to the noise continuum to remove the noise signal leaving only the detection data (the algorithm described in Starck et al., 2002). Duchamp allows the cube inversion and search for the negative detection, which could be used for the search of absorption lines or evaluating the reliability of the detections through negative source detection.

For the search in ALMACAL-CO data, we tested a set of parameters available in Duchamp to find the optimal settings: primary search mode spectral or spatial, the type of noise handling, wavelet reconstruction and smoothing, as well as the limiting signal-to-noise ratio. There was no significant difference between the results of the search between the spatial and spectral as the primary search axis. After the set of tests on the cubes with implemented mock sources, we chose the best performing set of parameters: wavelet reconstruction of the order of three with a signal-to-noise detection threshold or 3σ and the primary search in the spectral axis. The wavelet reconstruction method, however, has long computation time: a search over the whole database of 150 cubes would take several days to a few weeks. In the process of adjusting the search parameters for the larger dataset, this solution is not optimal.

Additionally, to the long processing time, Duchamp uses the global noise properties of the datacube for the candidate's selection. That approach did not work for the ALMACAL datacubes for which the noise can vary significantly between channels. The resulting list of candidates ranged from a few to hundreds, requiring additional filtering of the candidates and reliability analysis.

SoFiA

The alternative to the Duchamp source finder is the currently developed new generation of source finder for HI sky emission search, SoFiA (Serra et al., 2015). Optimised for searching through 3D data cubes, SoFiA offers a wide variety of search parameter. Similarly to Duchamp, SoFiA performs the search in the spectral and spatial dimension and merges the pixels with the detections in the final steps of providing the candidates' catalogue.

	Duchamp	**SoFiA**
Search algorithms	Spectral smoothing **Wavelet reconstruction**	**Spectral and spatial smoothing with adjustable kernels** Wavelet reconstruction and others
Speed	Wavelet: slow, single cube around 30 min - 1h	Smoothing: fast, single cube under 3-5 minutes
Number of detections	Large catalogues of sources, (20-100) requires additional filtering	Low number of sources (1-30), **Provides reliability parameter**
Noise handling	Global noise properties	Adjustable, global or local noise properties
Development	Not upgraded since 2014	Constantly upgraded

Table 3.5: A comparison of the properties of two source finders used in the ALMACAL-CO untargeted search for CO emission lines.

It has expanded features, enabling the user to choose the type of search, noise handling or reliability evaluation of the candidates. Within the source finder, we can indicate the region to be masked as well as flagged channels which should not be included in the search. Available search algorithms assume different treatment of the noise and continuum background. As in Duchamp, the main algorithms are smooth and sigma clipping, and wavelet reconstruction. SoFiA recommends the use of sigma clipping with smoothing as it adapts for a noise varying across the channels as is the case for the ALMACAL-CO sample. This feature is missing in the Duchamp source finder.

The smoothing and sigma clipping search operates through the set of different kernels varying in spectral and spatial size. The big advantage of this method is its speed, each cube can be analysed in under 3–5 minutes. The spectral smoothing approach allows also the choice of the smoothing kernel shape to be Gaussian or boxcar, making the search sensitive to the different types of detection.

The unique function of SoFiA is the noise properties calculation. For the detection threshold, we can use the global estimations of r.m.s. or define it separately for individual channels. The second option proved to be the best way to solve the problem of multiple detections which plagued the Duchamp searches. Additionally, one can choose what kind of noise statistic calculations is preferred: standard deviation, median absolute deviation, through fitting a Gaussian to a flux histogram or fitting a Gaussian function to the negative part of the flux histogram.

Lastly, SoFiA has an included reliability calculation method based on estimations of the negative detection fraction (Serra et al., 2012). Assuming all negative sources to be

artefacts, which in case of our data set is a valid assumption (we do not expect any absorption features in this data as the bright background calibrator is removed), the reliability of the positive sources is calculated from the comparison of the statistics of positive and negative detections (like a signal to noise ratio form the whole detection, from the peak flux etc). Positive detections in the region on the phase space populated mostly by negative detections are most likely also noise peaks. The typically recommended reliability value applied in SoFiA is 90 per cent.

SoFiA provides a range of the diagnostic files, including a cube with masks over the candidates' positions, their size in the spectral and spatial direction, an ASCII catalogue and a VOTable, including the detection parameters as well as flux maps of each detection. Additionally, for each candidate, an accompanying cubelet centred on the detection with the integrated spectrum is produced.

3.4 Volume calculation

To derive a reliable measure of the evolution of the cosmic molecular gas mass density $\rho(\mathrm{H_2})$, and hence the $\mathrm{H_2}$ luminosity function at different redshifts, we first need to evaluate the comoving volume probed by our survey. This volume needs to be calculated as a function of redshift and CO luminosity, separately for each CO transition that is probed by the survey. This can be achieved by first determining comoving volume elements dV (a volume measure in which number densities of non-evolving objects located into Hubble flow are constant with z), corresponding to redshift element dz and solid angle element $d\Omega$:

$$dV_c = D_H \frac{(1+z)^2 D_A^2}{E(z)} d\Omega dz, \tag{3.3}$$

where D_H is the Hubble distance ($D_H = \frac{c}{H_0} = 3000h^{-1}$ Mpc), D_A is angular diameter distance at redshift z and $E(z)$ is the cosmological function defined as:

$$E(z) = \sqrt{\Omega_\mathrm{M}(1+z)^3 + \Omega_\mathrm{K}(1+z)^2 + \Omega_\Lambda}. \tag{3.4}$$

For the Planck cosmology (Planck Collaboration et al., 2016), $\Omega_\mathrm{K} = 0$ (flat Universe), $\Omega_\mathrm{M} = 0.315 \pm 0.007$ and $\Omega_\Lambda = 1 - \Omega_\mathrm{M} - \Omega_\mathrm{K}$ (since all omegas summed together should equal to 1).

An extra complication arises because the sensitivity of an interferometer such as ALMA is not uniform across the field-of-view, but instead drops off gradually from the centre of the primary beam. In the calculations, I assumed that the sensitivity profile can be modeled with a Gaussian function and I assigned the sensitivity fraction of one in the centre ($r = 0$ in cylindrical coordinates) and the model-dependent value at r_max equivalent to the adapted image size of the datacube. To translate the sizes on the sky to the cosmological values we multiply the r_max by the scaling factor p, dependent on the redshift (in practice scaling

Transition	Redshift	Volume [cMpc^3]
CO(1-0)	0.00 - 0.35	910
CO(2-1)	0.00 - 1.70	17620
CO(3-2)	0.00 - 3.15	36025
CO(4-3)	0.10 - 4.40	46500
CO(5-4)	0.20 - 5.80	54840
CO(6-5)	0.01 - 7.10	60570
total	0.00 - 7.10	216470

Table 3.6: Volume per CO transition covered by the ALMACAL-CO sample. Depending on the frequency coverage each transition is traced in a different redshift bin. The dependence of the volume on the CO luminosity is shown in Figure 3.13. Columns: (1) covered CO transitions, (2) - redshift coverage of the transition, (3) Comoving volume corresponding to the transition.

through the cosmological angular distance) :

$$R[\mathrm{kpc}] = p[\mathrm{kpc/arcmin}] \times r[\mathrm{arcmin}] \tag{3.5}$$

In the sensitivity-limited case, the total cosmic volume accessible to a hypothetical galaxy of a certain luminosity would therefore have an "ice cream cone" shape for each individual calibrator field. To estimate the lowest detectable luminosity I used a modelled emission line with a boxcar shape, a width of $v = 200$ km s^{-1} and maximum flux (F) of 5 $\times$ the r.m.s of the cube:

$$L_{\mathrm{line}} \; = \; 1.04 \times 10^{-3} \, F \Delta v D_L^2 \nu_{\mathrm{obs}} \, [L_\odot], \tag{3.6}$$

and converted these parameters to L'_{line} [K km s^{-1} pc^2] luminosity following Carilli & Walter (2013):

$$L_{\mathrm{line}} \; = \; 3 \times 10^{-11} \nu_{rest}^3 L'_{\mathrm{line}} \tag{3.7}$$

To keep the emission line signal unresolved, I assigned it the spatial size of the synthesized beam.

I expressed the limits of the volume integration as a function of the luminosity (L'_{line}), however, the $\mathrm{D_L}$ and z are entangled in the L'_{line} definition:

$$\frac{D_{\mathrm{L}}^2}{(1+z)^3} = \frac{L'_{line} \nu_{obs}^2}{3.47 \times 10^8 F \Delta v}, \tag{3.8}$$

and therefore cannot be solved in an analytical way. I used two 6th order polynomials (the coefficients of the best fit polynomial change at the $z \sim 1.5$) to approximate the $\frac{D_L}{(1+z)^2}$ expression.

In addition to the sensitivity limit, the individual volumes for each calibrator field are characterised by frequency limits at a high and low-redshift end. The high-redshift end of each volume accessible to a galaxy of a certain luminosity could, therefore, be defined by either the low-frequency edge of the observation, a sensitivity limit that drops from the centre of the field or a combination of both. Hence, for each field, we integrate volume elements dz along with rings from the centre to the edge of the field, for each ring integrating in redshift from the lowest value defined by the high-frequency edge to the highest redshift.

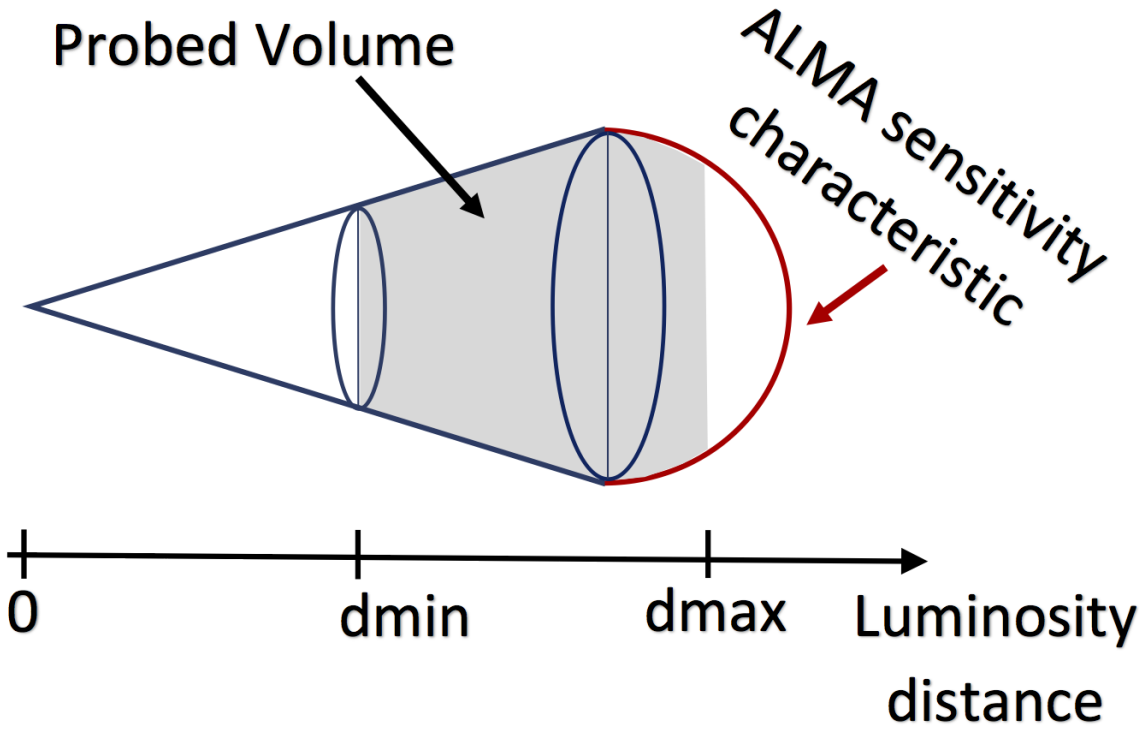

Figure 3.12: The sketch of the volume integration cone, including the *cap*(in red) from the varying sensitivity across the ALMA primary beam. The shaded area marks a volume element for integration within $d_{\rm min}$ and $d_{\rm max}$ limits. Integration limits are expressed as the luminosity distance and are calculated from the limiting line luminosity defined by the lowest accessible detection sensitivity. Integration limits differ for different CO transitions, line luminosities and can also be defined by the accessible redshift limited by the frequency probed by the cube.

We calculate these volumes for each field individually, each for a range of line luminosities from 10^5 to 10^{13} K km s^{-1}pc^2. Finally, we sum all volumes of individual fields to determine the total survey volume as a function of luminosity.

These comoving volumes are determined for each CO transition from CO(1−0) to CO(6−5), summing the volume coverage of each of the cubes for redshift from 0 to 5. The

results are presented in Table 3.6.

In Figure 3.13 we present the volumes per CO transition in different redshift bins as a function of the line luminosity. The sample is dominated by the high-frequency datacubes (Band 6, 7, 8) and we probe larger volumes for high-J transitions. The frequency coverage is complementary to the work of previous untargeted surveys (Decarli et al., 2019; Riechers et al., 2019; Lenkić et al., 2020), mapping the different CO transitions and redshift parameter space. Additionally, we scale the volumes probed by the pilot to the full ALMACAL sample which will be analysed in the future. The black solid line on Figure 3.13 indicates a factor of 50 increase of the total volume ALMACAL-CO survey per each CO transition with respect to the pilot survey presented in this work.

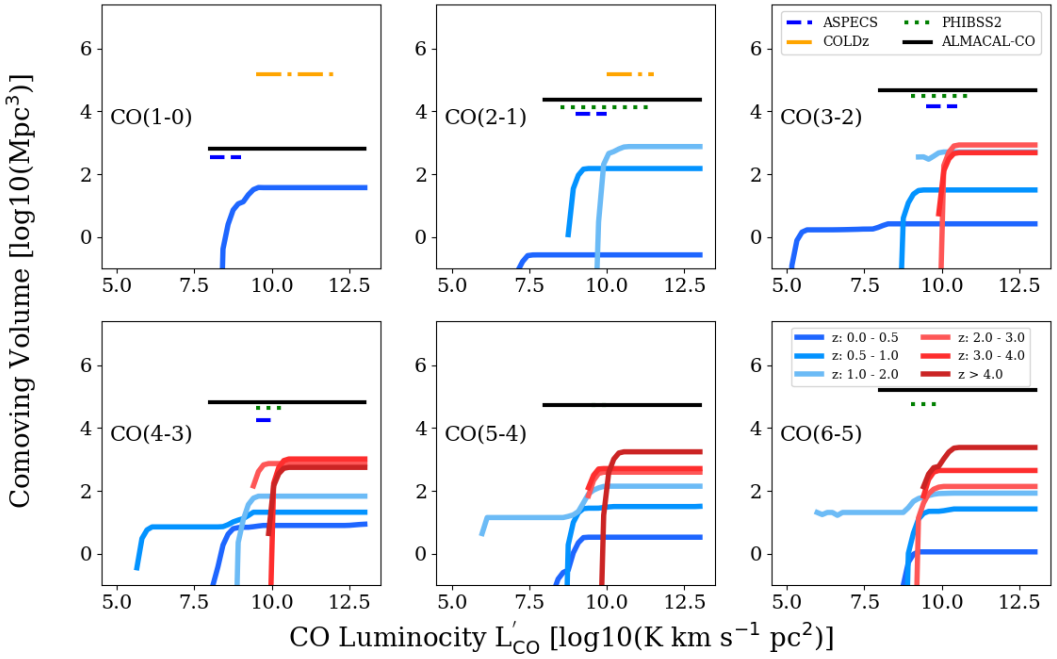

Figure 3.13: ALMACAL-CO pilot volume coverage for a range of line luminosities of different CO transitions (from J_{up}=1 to J_{up}=6). The range of lines from blue to red represents different redshift bins as indicated in the lower right panel. Dashed and dotted lines mark corresponding volumes probed by other untargeted emission-line surveys (ASPECS - yellow (Decarli et al., 2019), PHIBSS-2 - green, (Lenkić et al., 2020), COLDz - dark blue (Riechers et al., 2019)). Black solid line indicates the estimated volume probed but the complete sample of ALMACAL-CO data. ALMACAL-CO pilot covers smaller volume for individual transitions due to the frequency sampling and depth but because of the large frequency span, we are sensitive to a wide range of CO luminosities and transitions. Integrated over all redshifts, total volume of our pilot survey is 2–5 times higher than previous surveys.

3.5 Summary

ALMACAL is an ongoing project using numerous ALMA calibration datasets, allowing for conducting wide surveys with ALMA observations. The wide frequency coverage and multi-field approach allow a variety of extragalactic surveys and is less affected by the cosmic variance that single field observations. On the contrary to other surveys, the ALMACAL database constantly grows with each new ALMA science cycle, providing more detailed frequency coverage and deeper observations of the calibration fields.‘

Mining a big database such as ALMACAL requires a set of techniques, maximising the full usage of available dataset. In this chapter, I summarized the whole chain of data processing of the ALMA interferometric data for the ALMACAL-CO untargeted CO emission line search project: from the archive level to the ready-to-use data products, further analysis and source finding the 3D datacubes.

One of the important elements of the data analysis for the CO emission line search in particular, is source finding. The calibrator data pose a challenge to the analysis as the noise properties vary significantly between channels. I tested two source finders optimised for radio interferometry data and adjusted them to the ALMA characteristics. In the end, I chose the source finder SoFiA as the best performing solution.

The ALMACAL-CO pilot survey was proof of concept for the untargeted emission line search of the data, which results are presented in Chapter 4. In the future, based on the methodology and expertise from this exercise, I plan to extend the emission line search to the whole ALMACAL database (see Summary Chapter, Future work for details.)

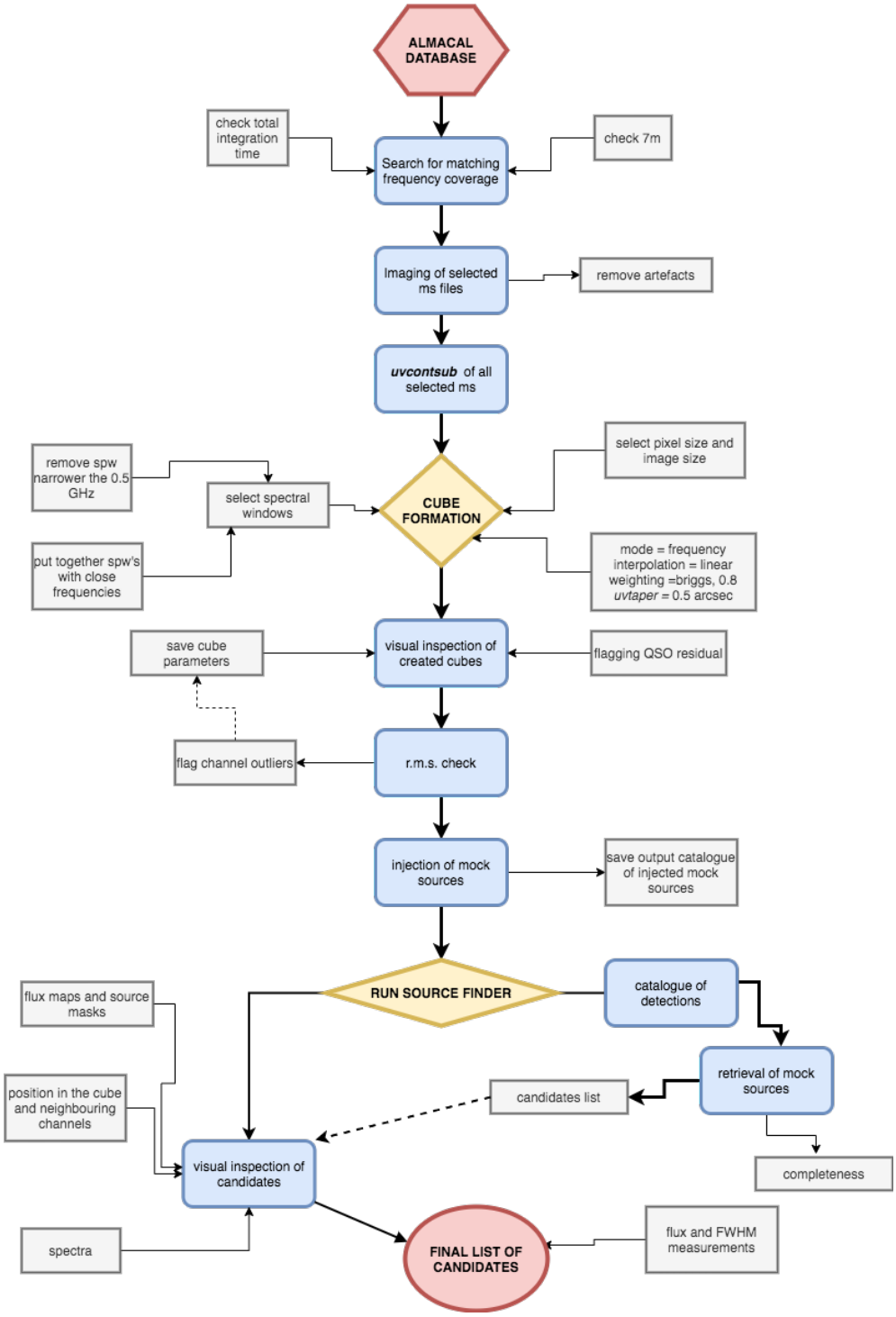

Figure 3.14: The chain of analysis processes from the ALMACAL data mining through the data reduction, source finding and final emission lines candidates catalogue. This set of procedures was applied to the ALMACAL-CO pilot project and will be used in ALMACAL-wide CO emission line search.

Chapter 4

ALMACAL-CO pilot, an untargeted CO emission-line search in deep ALMA calibration data

The content of this chapter is presented in the article Hamanowicz et al. to be submitted to the MNRAS.

4.1 Introduction

The cosmic star formation history (SFH, Madau & Dickinson, 2014) establishes the peak of the star formation in the Universe two billion years after the Big Bang ($z \sim 2$) followed by an order of magnitude decline to the present day. To understand what drives the cosmic star formation history we need to look at the constituents of the star formation, regulated by the evolution of cosmic baryons. The most abundant element in the Universe, hydrogen, is the main constituent of the stars and galaxies and, in its molecular phase, is the direct fuel of the star formation (Kennicutt & Evans, 2012; Bonato et al., 2018). To understand the cycle of baryons the global abundances of different phases of hydrogen are studied and compared to the evolution of star formation history (e.g. Péroux & Howk, 2020).

The largest reservoir of the gas available for star formation is hydrogen in its neutral phase. Although not a direct fuel of star formation, replenishing of the neutral HI by galaxies is vital process sustaining their star formation potential. Availability of the neutral HI gas thought the cosmic history influences the star formation processes on the scales of galaxies but could be also globally compared to SFH through measurements of neutral gas mass density $\Omega_{\rm HI}$. $\Omega_{\rm HI}$ are obtained mostly through two methods: 21-cm emission at radio wavelengths and intervening Lyα absorption features in distant quasar spectra (Rhee et al., 2018, and references therein). The 21-cm HI emission is directly measurable only in the nearby Universe, up to $z \leq 0.2$ (e.g. Zwaan et al., 2005; Jones et al., 2018), which could be extended to redshifts $z < 0.4$ trough stacking experiments (e.g. Lah et al., 2007; Rhee et al., 2013; Hu et al., 2019). Higher redshift measurements of $\Omega_{\rm HI}$ relies on the hydrogen

absorbers observed in the spectra of distant quasars as the sensitivity of these observations is independent of redshift and depends only on the brightness of the background source (e.g. Zafar et al., 2013; Rao et al., 2017). Mapped by these extensive observations, the $\Omega_{\rm HI}$ as a function of redshift is relatively flat between $z = 0$ and $z \sim 4$ and does not resemble the peak and decline of the star formation history of the Universe. To understand the evolution of SFH we have to study more direct probes.

Neutral hydrogen HI is not directly related to the Star Formation as stars form in the dense molecular clouds dominated by molecular hydrogen H_2. Variation in the availability of that molecular fuel is another proposed driver of the evolution of SFH of the Universe. Direct measurements of the molecular gas mass density $\Omega_{\rm H_2}$ through H_2 emission are, however, impossible for extragalactic sources as this molecule has faint vibrational transitions (Leroy et al., 2008; Schruba et al., 2011). Instead, the molecular gas content of galaxies relies on the observations of the second most abundant molecule, CO, whose bright transitions are observable also at high redshifts ($z > 2$ Carilli & Walter, 2013; Hodge & da Cunha, 2020). CO has been detected in star-forming galaxies at different redshifts through targeted surveys (e.g. Daddi et al., 2010; Genzel et al., 2010; Freundlich et al., 2019; Tacconi et al., 2020), providing us with the view of the molecular gas reservoirs in massive galaxies and their link to star formation. Such surveys are targeting massive star-forming galaxies and establish the scaling relations linking the galaxies' molecular gas content with their other properties like stellar mass or star formation rate. Extrapolating these scaling relations to lower mass regimes provides an approximate overview of the molecular gas content of galaxies at different redshifts.

Although powerful, scaling relations from targeted surveys are introducing selection biases influencing the measurements of $\Omega_{\rm H_2}$, which cannot be quantified in detail. An unbiased way to study the molecular content of galaxies across cosmic time is through untargeted emission-line surveys: observation over a selected sky area without target preselection. Untargeted CO surveys have so far focused on cosmological fields that have a significant multi-wavelength coverage: ASPECS (Walter et al., 2016; Decarli et al., 2019) in HUDF, COLDz in GOODS fields (Riechers et al., 2019) or through mining the ALMA archival observations of the COSMOS field (Liu et al., 2019). These surveys have been limited to single fields, covering a small part of the sky, which raises concerns about the possible effects of cosmic variance. Additionally, in a more universal approach, (Lenkić et al., 2020) searched for secondary sources in the NOEMA PHIBSS-2 survey observations (originally survey targeting CO emission in massive star-forming galaxies). To quantify the molecular gas budget of the Universe and the impact of the molecular gas content of galaxies on the Star Formation History of the Universe (SFH) we need a statistically significant sample of CO-bright galaxies probing a range of stellar masses. In this work, we present the ALMACAL-CO pilot survey: an untargeted search for CO emission in ALMA calibration data obtained semi-randomly across the southern sky. This pilot survey is proof of concept for using the ALMA calibrator data to perform a more comprehensive untargeted emission line search. We are also introducing a novel statistical approach, addressing the emission line classification challenges of a sky-wide survey with the shortage of corresponding deep optical follow-up observations. Instead of search for optical counterparts for each of the

detections, we construct the redshift probability functions of each emission line based on the CO flux predictions from the Shark Semi-Analytical Models (Lagos et al., 2018). The results presented in this chapter are preliminary; with the future main survey spanning over all ALMA calibrator fields, we will significantly improve the statistics and be able to provide better constraints on the CO luminosity functions as well as the evolution of the molecular gas mass density with redshift.

4.2 An untargeted emission-line survey

4.2.1 The ALMACAL survey

The ALMACAL project[1] is utilising the archival observations of ALMA calibrators for scientific purposes (Oteo et al., 2016). Each scientific observation with radio interferometry requires several short integrations on a calibrator located at a small separation from the science field. The calibrator is observed with short exposure time (several minutes) with the identical setup as science observations requested by the project's PI (spectral resolution and coverage). In addition, less frequent, but deeper integrations of bandpass calibrator are taken. ALMA calibrators are bright sub-millimetre point sources, distributed over the sky accessible to the observatory. Ninety-seven per cent of these sources are classified as blazars, and 78 per cent have known redshift (Bonato et al., 2018).

Calibration scans become publicly available after the data sets that include these scans have passed quality assurance. For every new observation, the calibrator scans are delivered together with the science data. We extract data out of ALMA archive, remove the science observations and reduce the calibration data separately, self-calibrating on the central bright source with a custom-made pipeline. All data are resampled to the same spectral resolution of 15.5 MHz and self-calibrated on the central source determining solution on the shortest time intervals. Finally, the bright point source is modelled and subtracted from the visibility data, leaving a calibrator-free data set (see Chapter 3 for detailed description). At the present stage of the ALMACAL project, the database consists of 880 calibrator fields. We are accumulating new data constantly, but only sporadically are new fields added. The frequency coverage and depth of the observations depend on the popularity of the calibrator. Example of the frequency coverage of the ALMACAL fields can be found in Figure 4.1 (for fields included in this study). The ALMACAL database includes observations from all ALMA bands, the majority of which are in Band 3 (84–116 GHz) and Band 6 (211–275 GHz). Up until today, the ALMACAL database consists of over 2000 hours of observations (the equivalent of one-third of the one full ALMA yearly observing Cycle) a number that is growing with every new science ALMA observations.

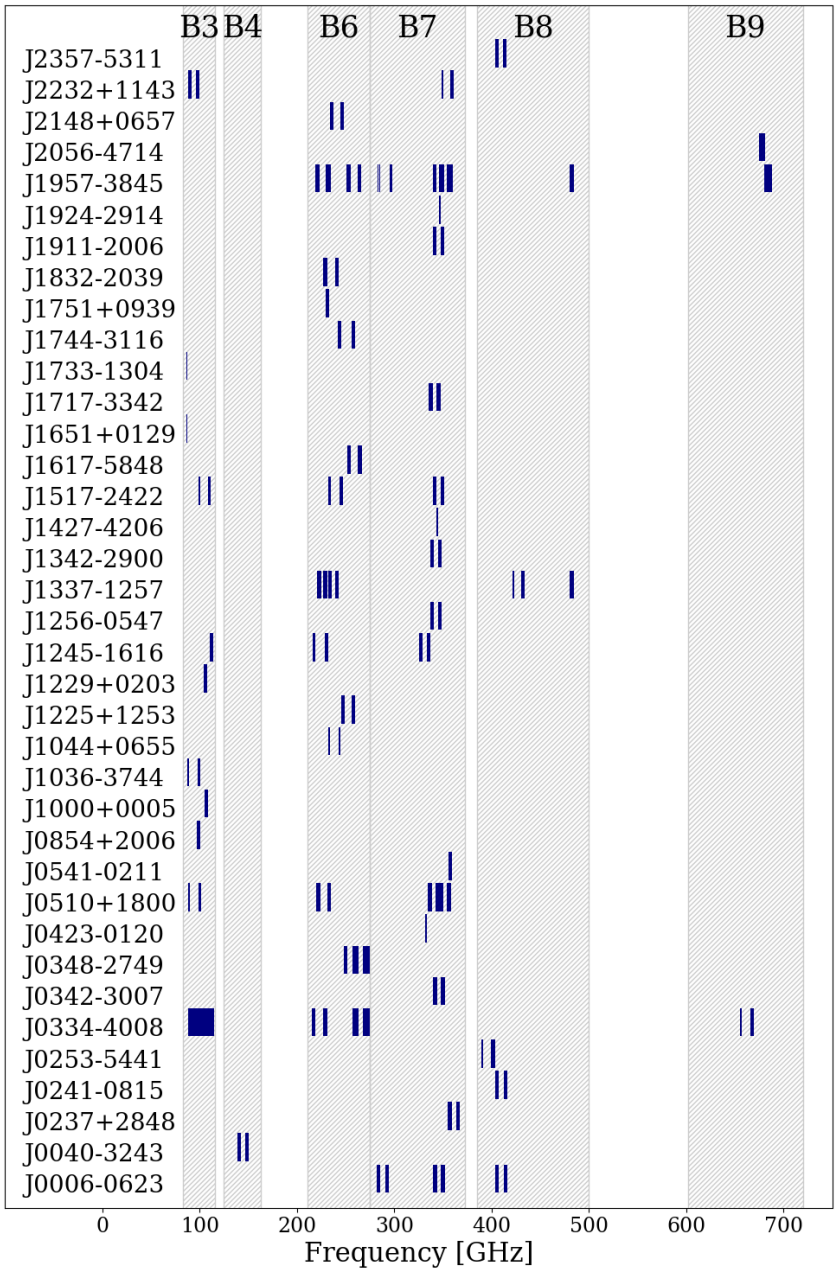

Figure 4.1: Frequency coverage of ALMACAL-CO deep fields ($T_{int} > 40$ min). The fields are presented with increasing right ascension from bottom to top. Blue rectangles mark the frequency coverage of the data used in this study. Grey shaded areas represent the ALMA bands' frequency coverage labelled on the top. The frequency coverage of ALMACAL observations depends on the PI's requests of corresponding ALMA science projects. The coverage of different calibrators depends on their popularity, mainly the position on the sky near the most often observed targets.

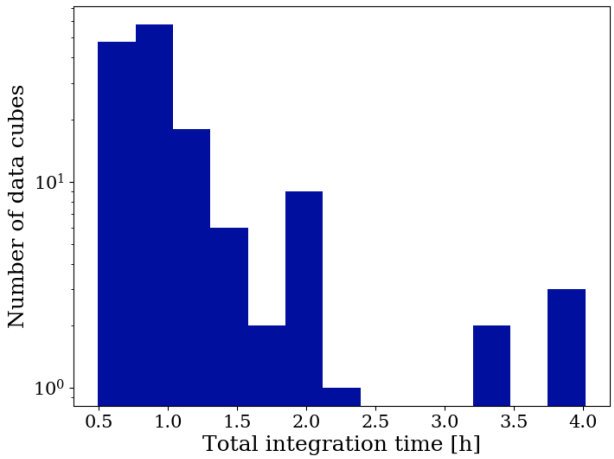

Figure 4.2: Total integration time distribution of the ALMA data-cubes used in this study. The lowest integration time is limited to 40 minutes by construction. The longest integration time reaches 2 hours with a few exceptions of even longer observations. The total integrated time results from concatenating the short (several minutes) singular calibration pointing taken in the same spectral setup.

4.2.2 ALMACAL-CO sample selection

ALMACAL-CO is a project aiming at measuring the evolution of the cosmic molecular gas mass function through the untargeted detections of CO emission lines. In this work we present the pilot project, focusing on the fields with the longest integration times, to which we further refer as *deep*. The data selection was performed in September 2017. Below we describe the selection of the data for the pilot study.

For each ALMACAL calibrator field, we chose observation undertaken in the same frequency setup accumulating a total observation time longer than 40 minutes. The cubes vary in depth (integrated time) from 40 minutes to 4 hours (Figure 4.2). For each field, the chosen observations were concatenated into a single file and reduced to a data cube with standard CASA routines. Before concatenation, we imaged each observation and checked the quality visually, to remove any data suffering from artefacts related to imperfect calibration or dynamic range limitation. The selection of the data resulted in 147 cubes over 37 calibration fields (the distribution of the cubes on the sky is presented in Figure 4.3), each cube representing a different frequency coverage. To remove potential residuals from the quasar subtraction we ran the *uvcontsub* routine with the fit order equal to one, on the concatenated files before imaging. The continuum was fitted in the frequency domain and any large scale noise structures were removed. Even after removing the continuum, some of the observations were contaminated by non-linear residuals of quasar subtraction, typically remnants from atmospheric corrections. In these cases, we masked the central

[1] `almacal.wordpress.com`

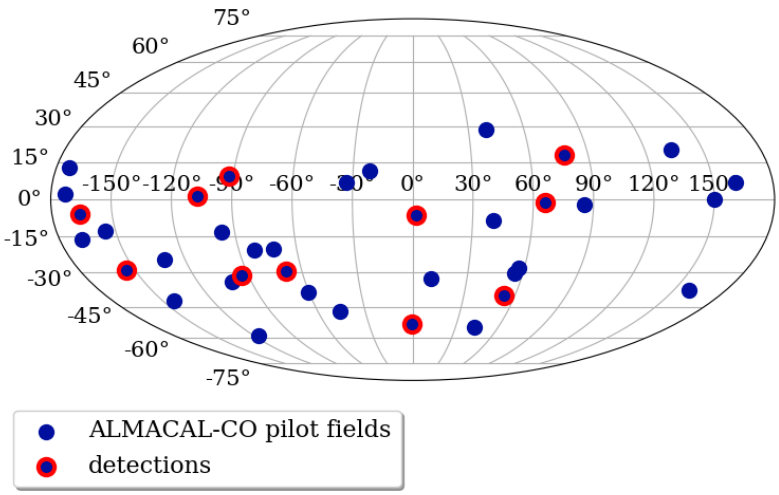

Figure 4.3: The ALMACAL-CO deep field distribution on the sky (blue circles). Red circles mark fields with reported detection included in this study. The size of the points does not represent the physical scale. The random distribution of fields over the sky makes the survey less prone to the effects of cosmic variance.

part of the cube (we used the squared mask with the equivalent of two beams).

The spatial extent of each image cube is determined by the primary beam size, depending on the observed frequency. It ranges from 56 arcsec in Band 3, 27 arcsec in Band 6 to 9 arcsec in Band 9. We used *CASA Analysis Utilities* package routines to estimate the best parameters for the construction of the images: the size of the pixels and the number of pixels across the image. In constructing the cubes, we combine data with a large range of spatial resolutions. In order to produce a dataset with consistent properties, which allow optimal detection of unresolved line emission, we taper all visibility data to the same spatial resolution of 0.5 arcsec. Furthermore, we limit the minimal pixel size to one-third of that value. We exclude the data with spectral windows narrower than 0.5 GHz. In this project, we also do not consider observations from the ACA array.

After the creation of the data cubes, we calculated the total root-mean-square (r.m.s) of each cube and created the noise-per-channel plots. The resulting analysis showed that some channels extend over the mean noise levels and revealed highly variable noise patterns thought the cube. We removed from the sample data cubes for which the r.m.s was higher than 5 mJy Beam^{-1}. Additionally, we flagged the outlier channels which noise level was exceeding over 1σ of the noise variation in the cube (less than 20 per cent of all data).

Because of the nature of the ALMACAL data, the spectral coverage of each observation depends on the spectral setting defined by the PI of the corresponding science project (Figure 4.1). The non-uniform spectral coverage of our fields results in variation of the

volume probed. The inhomogeneity of the frequency coverage is taken into account in our volume calculations. In some of the calibrator fields, the frequency coverage is scarce, making the untargeted detection of the CO relatively unlikely.

4.2.3 Shark Semi-Analytical Models

Due to the limited spectral coverage per field, the probability of detecting more than one transition from the same objects is low. Additionally, deep ancillary data, which would ease the identification of the detections through optical counterparts, are mostly not available for the calibrator fields. To classify our candidates, we instead employ a probabilistic approach based on the catalogue of properties of galaxies simulated with the newly developed open-source, highly complex detailed physics code Semi-Analytical Models (SAMs) Shark (Lagos et al., 2018, 2019). Ideally, instead of SAMs, for the probabilistic exercise, we would use the observed database of CO emission lies. However, currently available observations of CO emission (especially multi-CO line observations from the same object) are not extensive enough to realise such an approach.

Shark SAMs are based on a hydrodynamical Dark Matter only simulation, which halos are populated with galaxies without resolving their inner structure. SAMs can generate large cosmological light-cones (the largest available from Shark covers ~ 108 deg^2) and trace the evolution of millions of objects through cosmic time. With these properties, these simulations are ideal tools for large scale sky surveys predictions and inter-match between expected and measured quantities including the general evolution of cosmic densities.

Shark is based on the *SURFS* n-body simulation of dark matter halos starting at $z = 24$ (Elahi et al., 2018), following the (Planck Collaboration et al., 2016) ΛCDM cosmology. The properties of simulated halos are then calculated using the *VELOCIRAPTOR* (Cañas et al., 2019) and the merger trees of the emerging halos are traced using the *TREEFROG* (Elahi et al., 2018). Based on the parameters of the dark matter halos and their merger history, Shark evolves galaxies in their halos in every time snapshot using the physical model described in (Lagos et al., 2018). The model describes the physical processes governing the evolution of galaxies including star formation, stellar and AGN feedback, gas accretion, gas shock heating and cooling, mergers and environmental effects, disk instabilities, chemical enrichment, photoionisation feedback and black hole growth. For the details of the prescriptions and codes used for the creation and evolution of an object in the simulations see (Lagos et al., 2018). Combining the models with the spectral energy distribution software tool *PROSPECT* (Robotham et al. in prep.) result in the generation of a SED for each simulated galaxy, spanning from FUV to FIR (Lagos et al., 2019) including dust attenuation. All galaxies, additionally to their molecular gas masses, have predicted CO luminosity SLEDs following the recipe from (Lagos et al., 2012). Multiple simulation cubes are rotated (to prevent artificial clustering) stacked and evolved over the broad redshift range creating a significant cosmic volume.

The SAMSs reproduce a wide range of galaxy properties and types including metal-poor systems and SGMs and reproduce the galaxy luminosity functions and the number counts of sub-millimetre galaxies. The results of the simulations compared with the observations

Table 4.1: ALMACAL-CO emission line candidates. The two detections in the field of calibrator J0334-4008 coincide with nearby galaxies detected with VLT/MUSE, allowing for the redshift identification of the sources as CO(1-0) lines. For these two detections we assign the probability function of 100 per cent for the spectroscopically confirmed transition. The columns: (1) Candidate ID, (2-3) sky coordinates of the detection in J2000, (4) central frequency of the detection in GHz, (5) integrated line flux, (6) width of the detection, (7) signal-to-noise as reported by the *SoFiA* source finder.

Source	RA(J2000)	DEC(J2000)	Frequency	Flux	FWHM	S/N
	[hh:mm:ss]	[dd:mm:ss]	[GHz]	[mJy km s^{-1}]	[km s^{-1}]	
J000614.36−062329.4	00:06:14.36	−06:23:29.4	348.46	50 ± 10	35 ± 5	5.9
J042316.48−012023.7	04:23:16.48	−01:20:23.7	330.64	65 ± 22	30 ± 8	4.9
J051002.35+180052.5	05:10:02.35	+18:00:52.5	344.57	80 ± 13	35 ± 4	5.1
J134214.60−290046.5	13:42:14.60	−29:00:46.5	344.36	70 ± 24	30 ± 3	4.1
J165104.43+012924.8	16:51:04.43	+01:29:24.8	346.79	55 ± 10	30 ± 14	4.4
J174423.10−311617.3	17:44:23.10	−31:16:17.3	231.93	90 ± 19	55 ± 9	6.7
J175133.98+093846.4	17:51:33.98	+09:38:46.4	242.79	45 ± 17	90 ± 26	5.1
J192450.65−291444.7	19:24:50.65	−29:14:44.7	233.48	40 ± 13	60 ± 15	4.9
J235752.46−531120.7	23:57:52.46	−53:11:20.7	135.91	60 ± 10	200 ± 26	6.7
J033416.50−400816.0	03:34:16.50	−40:08:16.0	97.23	100 ± 22	255 ± 43	5.9
J033412.22−400806.8	03:34:12.22	−40:08:06.8	101.67	40 ± 7	60 ± 9	6.4

recreate well the cosmic star formation date density at different redshifts (assuming a higher H_2-to-stars conversion at $z \leq 1.5$) as well as galaxy stellar mass functions.

4.3 Untargeted CO emitters counts

4.3.1 Detection of Sources

The untargeted emission line search over ALMA datacubes in the ALMACAL-CO pilot sample was done using the *SoFiA* open-access source finder (Serra et al., 2015). *SoFiA* is a flexible source finder developed primarily for searching and characterizing detections from large HI surveys. The source finder searches the 3D data cubes on multiple scales, both spectrally and spatially, allows for the choice of various source-finding algorithms including classical sigma clipping or more advanced wavelet reconstruction. It allows for the customised treatment of the noise, depending on the cube parameters, it can be calculated globally or in each channel separately. *SoFiA* estimates the reliability of the sources by measuring the negative detections in the inverse cube and comparing their properties with the positive detections. This way the user decides about the level of reliability allowed for the final catalogue of sources. We adopted a reliability cut of 90 per cent (R = 0.9 in Equation 4.1).

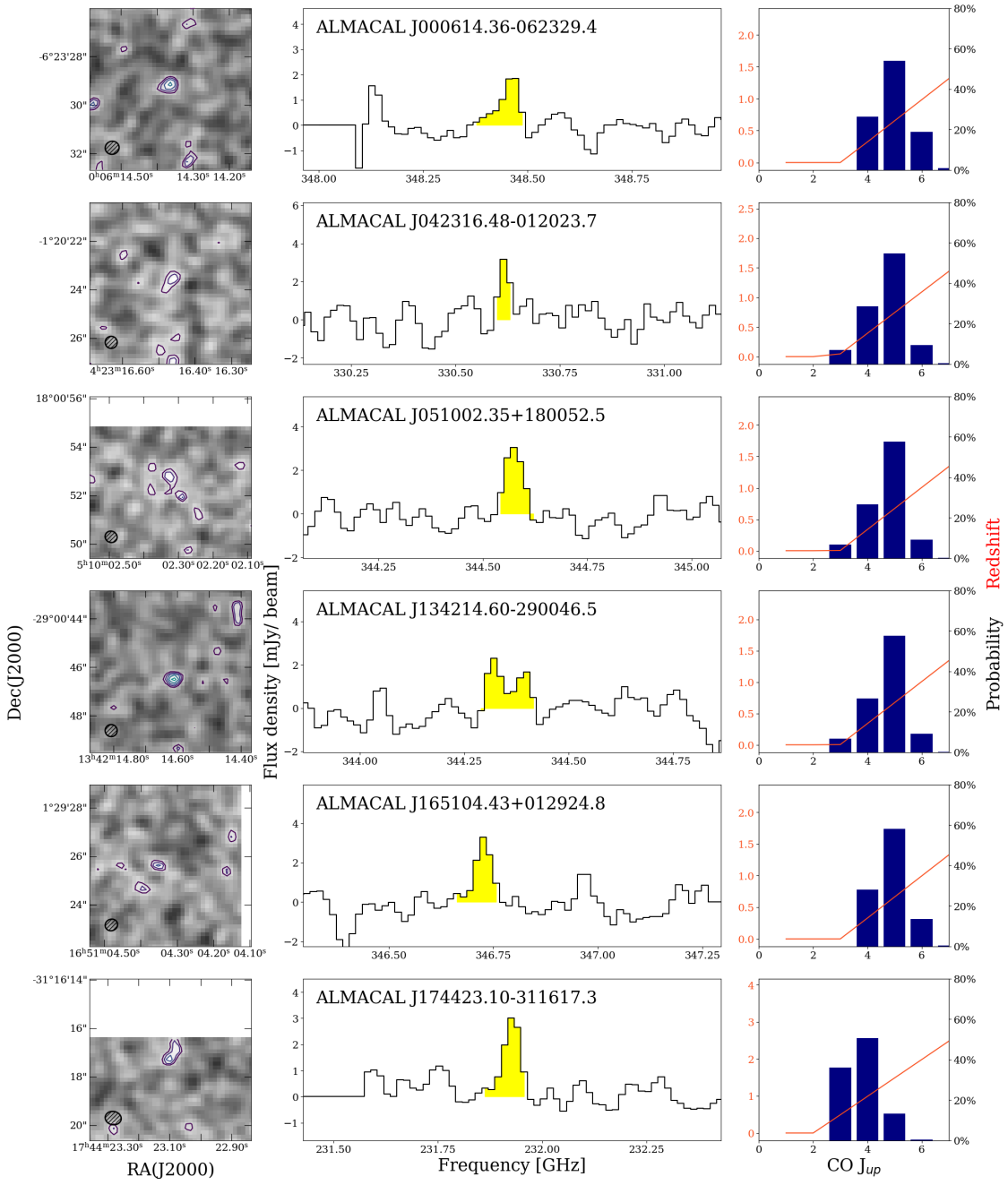

Figure 4.4: ALMCAL-CO CO emission lines candidates (see the next page for the description).

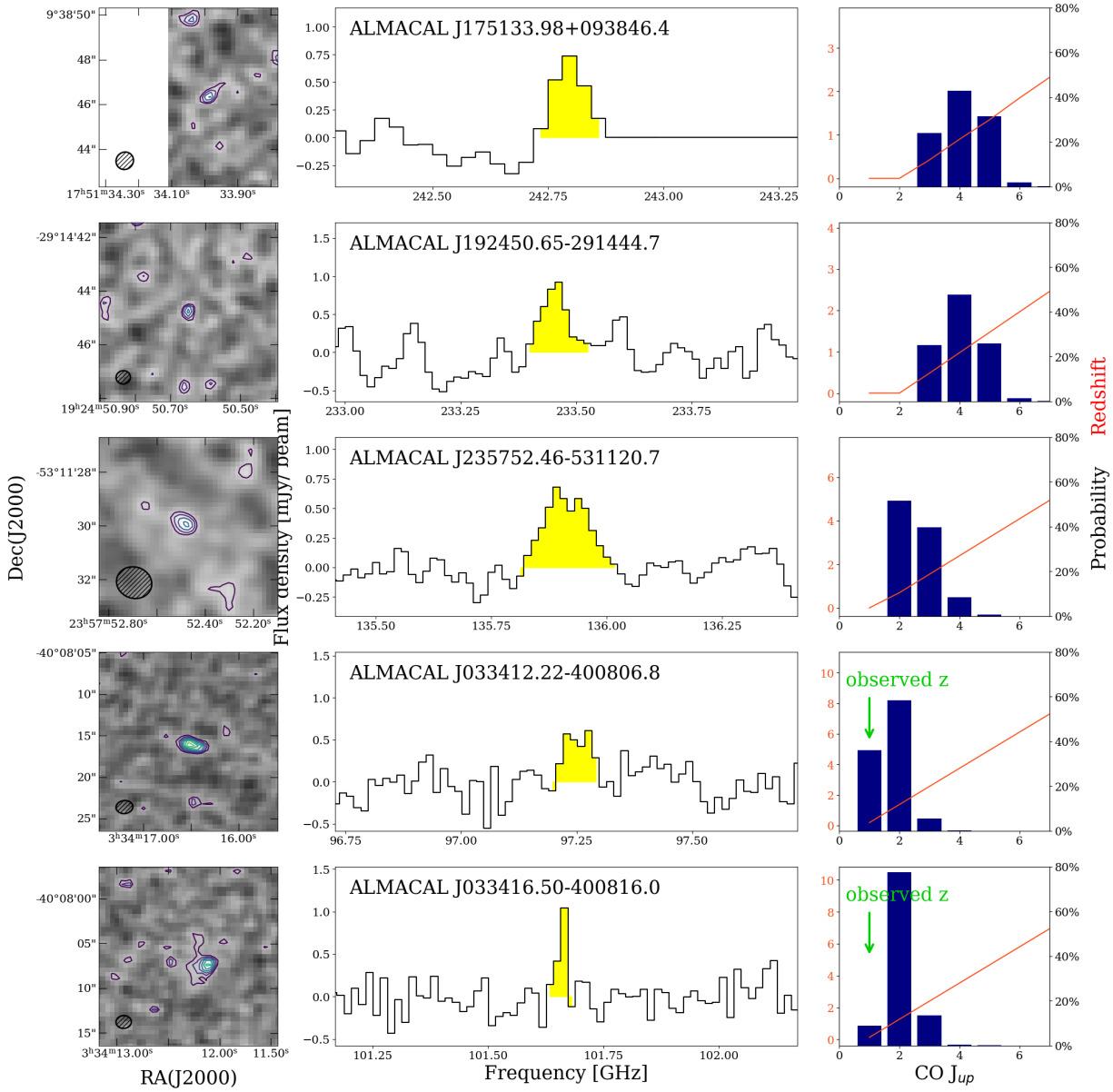

Figure 4.4: ALMCAL-CO CO emission lines candidates. Left: ALMA 0th-moment map centred on the detection candidates discovered in the ALMACAL-CO untargeted search. The contours correspond to +2,2.5,3,3.5,4...σ, derived from the respective line map. The size of the synthesized beam is shown in the lower-left corner. Middle: spectrum of the line candidates brightest pixel. Right: The distribution of CO transition predicted by Shark corresponding to ALMACAL-CO line detections. The majority of sources in our sample are associated with $J_{\rm up} = 4,5,6$. The parameters of all candidates as summarised in Table 4.1. For the candidates form field J0334-4008 (last two panels), we choose the CO(1−0) solution with probability 100 per cent, as confirmed by the optical counterparts.

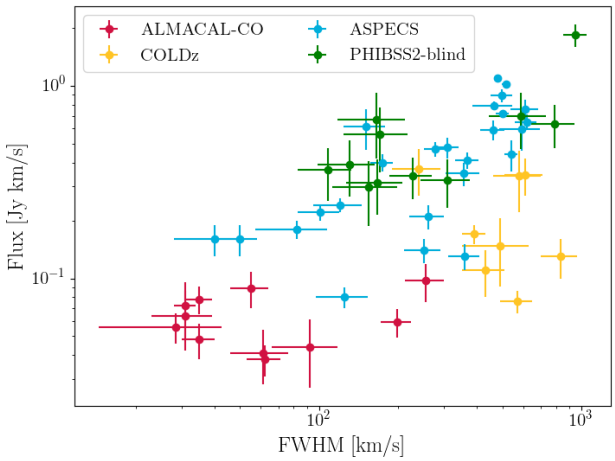

Figure 4.5: The integrated flux and width of the CO candidates from ALMACAL-CO (red) compared to detections from other untargeted surveys: ASPECS (blue, Decarli et al., 2019), COLDz Riechers et al. (yellow, 2019), PHIBBS-2 (green, Lenkić et al., 2020). ALMACAL-CO detections are fainter and narrower compared to the detections from other surveys.

The source finding was performed using the sigma clipping algorithm together with spectral smoothing over the set of kernels increasing spatially and spectrally. This method allowed for the fast analysis of the cube in only a few minutes. We limited the search to a 3σ threshold and required the minimal size of the detection to be two pixels and three channels. After the search, we identified the candidates and we further examined the spectra and flux maps of the detections visually. Judging by the quality of the spectra, noise properties or spatial profile of the detection we excluded about half of the original detections resulting in eleven final candidates. The properties of the line candidates are summarised in Table 4.1, while spectra and moment maps are shown in Figure 4.4.

The ALMACAL-CO survey is complementary to other untargeted CO emission line surveys conducted so far. We are probing higher frequencies thus are sensitive to high-J CO transitions around redshift $z = 1$–2. Our detections (Figure 4.5) are fainter and narrower than the ones reported by ASPECS and other surveys.

4.3.2 Estimation of redshift of the CO candidates

The challenge of using ALMACAL data for untargeted CO emission lines search is the uneven spectral coverage of each field. All our detected candidates are single-line detections, creating a degeneracy between redshift estimation and transition classification.

We used Shark Semi-Analytical Models (Section 4.2.3) of galaxy evolution to create the *redshift probability calculator*, returning redshift probability function for each detection, describing the likelihood for the candidate to be a given CO transition. We generated a

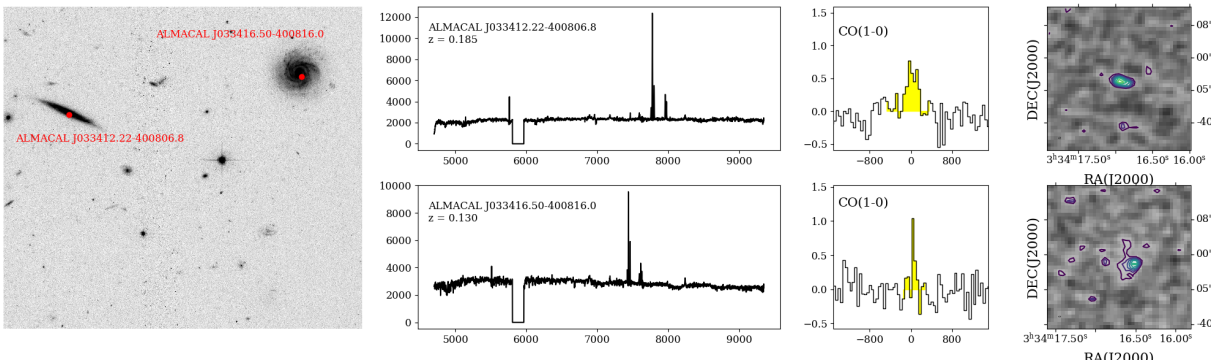

Figure 4.6: Combination of optical (HST imaging, MUSE spectroscopy) and ALMA observations of the ALMACAL-CO field J0334-4008. Two detected emission lines in that field coincide spatially with the two spiral galaxies, edge-on at $z = 0.185$ and face-on at z=0.133, allowing for the direct identification of the lines. The left panel shows the HST image of the field with the positions of ALMACAL-CO detections marked with red points. Second panes show MUSE spectra of the host galaxies, next panel presents the CO(1-0) detections for each galaxy accordingly. The most right panes show flux maps centred on the ALMA detections with +2,2.5,3,3.5,4...σ contours. J0334-4008 field detections are the only one in ALMACAL-CO sample with optical counterparts covered by archival data.

catalogue of Shark simulated galaxies, requesting the brightest CO transition to be brighter than 0.05 mJy km s^{-1}. The selection resulted in a catalogue of about 4 million sources covering redshifts from z = 0–6. For each CO transition, we created a 2D histogram binning the data of CO transition flux and redshift. For each bin of transitions $J_{\rm up} > 1$ we additionally save the CO(1−0) flux corresponding to each bin. Lastly, we divide the number of objects in each bin by the total number of objects in the simulated sample defining the probability coefficient for a certain pair of flux and redshift.

Because of the frequency coverage, bright low-J transition lines (especially $J_{\rm up}$ = 1, 2) are not available. On the other hand, high-level transitions ($J_{\rm up}$ = 8, 9, 10) are expected to be faint and their probability is close to or exactly zero. From the Shark probability functions, the most likely transitions corresponding to the detections are $J_{\rm up}$ = 4, 5, 6 which places most of our candidates between redshift 0.5 and 2.

4.3.3 Search for optical counterparts

The probabilistic approach, based on the Shark Semi-Analytical-Models, provides the population-weighted predictions for the observed lines. For a fully secure classification of detections, optical counterparts would be preferable. We searched several optical data archives for observations of the calibrator fields.

We expect the optical counterparts of the CO line detections to be similar, rather faint ($\sim$ 23–25 mag) sources as ASPECS counterparts (Boogaard et al., 2019). We checked the archives of optical all-sky Pan-STARSS1 (Chambers et al., 2016) survey, for a presence of

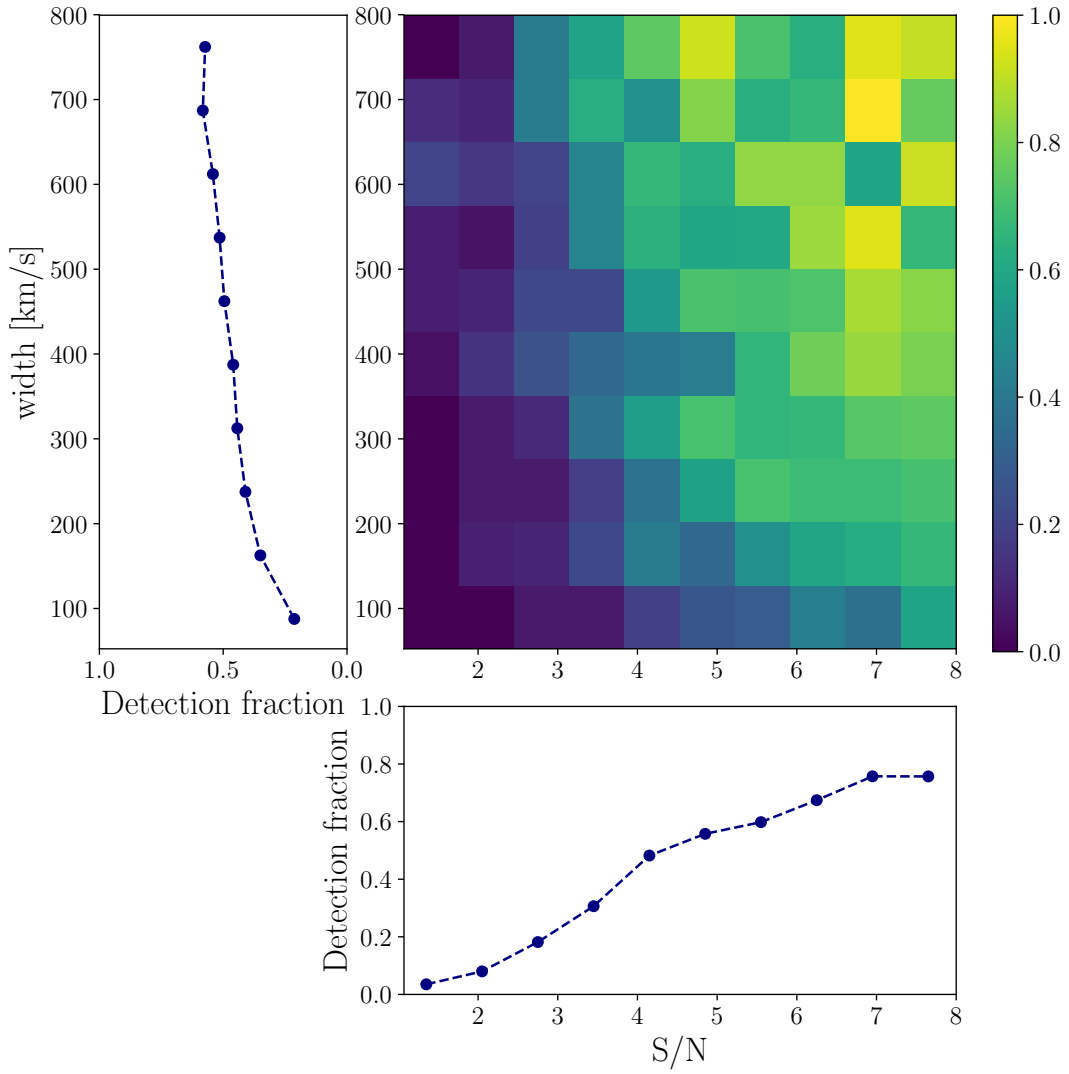

Figure 4.7: Completeness of the *SoFiA* search for line emitters in ALMACAL-CO pilot cubes. The detection fraction of line emitters as a function of their signal-to-noise ratio as well as the signal width in km s^{-1} is represented by colours on the 2D histogram. The additional panels represent the detection fraction as a function of one of the emission line parameters. The gradient in retrieving the unresolved mock sources points toward strong dependence of the success on the emission line velocity width as opposed to signal-to-noise. The highest completion rate has broad and strong signals (top right corner of the plot).

any continuum detection at the position of ALMACAL-CO line candidates. We did not detect any optical source at any candidate's position. With exposure times of the order of a few minutes, Pan-STARRS1 observations might not be deep enough to detect the counterparts of our sources.

Archival MUSE data were available for two fields: J0423-0120 and J0334-4008. The J0423-0120 field was observed for $T_{\rm exp}$= 1.5 h but no corresponding detection was found. For comparison, the ASPECS detections were reported in MUSE $T_{\rm exp}$ = 10 h observations.

In the J0334-4008 field, the two ALMACAL detections coincide spatially with two bright galaxies at redshifts $z_1 = 0.133$ and $z_2 = 0.185$ (see Figure 4.6), placing the line candidates as CO(1−0) transition at these redshifts. Those systems are the only sources in our sample with a confirmed redshifts and for the further analysis, we adopt that redshift solution with 100 per cent probability. We note that in the redshift probability function, the solution with $J_{\rm up} = 1$ was indeed assigned a significant probability (the highest was CO(2−1) at $z \sim 1.2$).

Additionally, we searched ALMACAL archives for the other transitions corresponding to the candidates, however, we do not report any detections.

4.3.4 Completeness

To measure the completeness of our untargeted search we inject mock sources into all cubes in the sample. We populate each cube with 20 mock sources and run the *SoFiA* source finder with the same set-up as for the actual procedure.

Mock sources are designed to resemble the possible real emission lines and are described by several parameters: shape, peak flux, central frequency and width. To test the recovery of the different emission lines shapes we assigned the mock sources randomly a gaussian, rectangle and double-horned shape (imitating the Doppler effect of a rotating object) with a random central frequency within the spectral range of a single cube. The peak flux was defined to have a range of 1 to 8 times the r.m.s of the noise in the given cube. The width of the detection was then assigned randomly from a uniform distribution between 50 and 800 km s^{-1}. Finally, we assign each detection a random spatial position in the cube, excluding a three-pixel thick frame around the edges. After each injection, the positions became forbidden for other ingestions within a ± 3-pixel radius (the equivalent of the beam size), to prevent the overlapping of sources. All mock sources are spatially unresolved.

We performed the source search with the *SoFiA* source finder, aiming to retrieve mock sources and detect real sources in the same search run. Afterwards, we compared the position of the detected sources with the mock sources catalogue and highlighted all objects not corresponding to any artificial injected source as a candidate.

The search completeness as a function of emission-line strength and width is shown in Figure 4.7. Following our expectations, we were most likely to detect broad and strong emission lines, for which the detection fraction reached 90–100 per cent. We are about 60–70 per cent complete for narrow emission lines. For emission lines below 3σ, we reach a lower level of completeness. The width of the emission line has a significant impact on

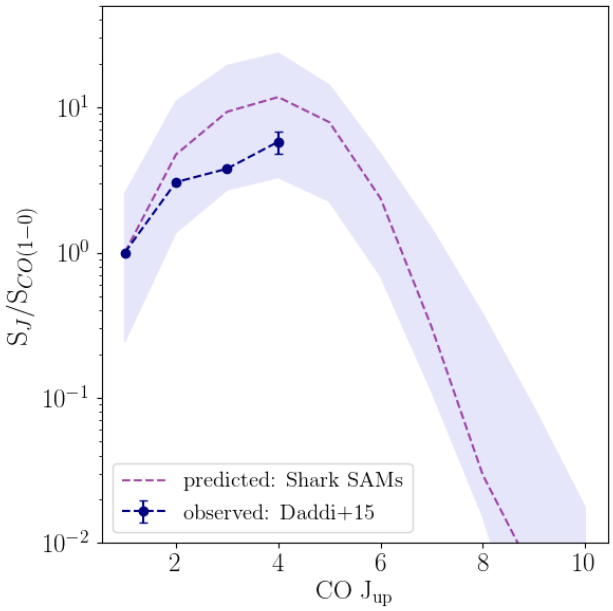

Figure 4.8: The average CO SLED of BzK $\langle z \rangle = 1.5$ galaxies simulated by Shark. The dashed purple line marks the mean SLED of the subset chosen to this test, the lavender area covers the 16th and 84th percentiles. The dark blue points and dashed line are a CO SLED averaged over four BzK galaxies at $\langle z \rangle = 1.5$ from Daddi et al. (2015). Te fluxes of all transitions are normalised to the CO(1−0) flux.

the detectability and for lines narrower than 200 km s^{-1} the completeness does not exceed 50 per cent even for the brightest sources.

We define the completeness coefficient as the fraction of retrieved sources for a given signal-to-noise and emission line width. The completeness fraction deduced from this analysis is used to correct the luminosity function and mass density function.

4.4 Comparison of Shark SAMs predictions with observations

4.4.1 CO SLEDs of Shark galaxies

The redshift estimations of the ALMACAL-CO detections are based on the CO emission line flux predictions from the Shark simulations. The relative strength of the CO lines in the CO ladder is described by the CO SLED (Spectral Line Energy Density) and depends on properties of the host galaxies, especially their ISM conditions. We want to test Shark CO SLEDs predictions against observations to see how well we are representing the observable galaxies. The SLEDs have a direct impact on the results of our survey as they determine the relative brightness of the high-J lines with respect to the CO(1−0). We base our redshift

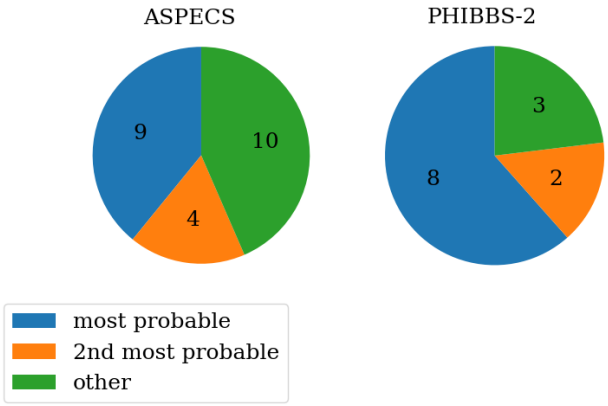

Figure 4.9: Number of correctly identified transitions of the sources from ASPECS (Decarli et al., 2019) and PHIBSS-2 (Lenkić et al., 2020) surveys by our Shark-based redshift calculator. The blue slice depicts the number of transitions for which the identification as most probable is consistent with the spectral classification. The orange slice covers transitions for which the most probable $J_{\rm up}$ was one lower/higher than the classified one (and the actually observed transition was assigned significant probability), while the green slice covers all other cases. The redshift calculator identifies correctly about 50 per cent of the detections (over 60 per cent if we include the "2nd most probable" option).

probability estimations on the fluxes of the emission lines as we compare the brightness of the observed line with simulated line fluxes of different CO transitions. Secondly, we converted observed CO luminosities to CO(1−0) transition (that is used for the molecular gas mass estimation) using the simulated values, which are also dependent on the CO SLEDs implemented in simulations.

We know only a handful of CO SLEDs in star-forming galaxies and some extreme objects like quasars, SMGs or starburst galaxies. The most extensive study so far are CO SLEDs from BzK galaxies at median redshift of $\langle z \rangle = 1.5$ in Daddi et al. (2010). The authors report the flux measurements for four CO transitions ($J_{\rm up} = 1,\ 2,\ 3,\ 4$) in four galaxies creating an average CO SLED which we used for the comparison with the SLEDs simulated by Shark. BzK galaxies represent the typical star-forming object in the Universe above $z > 1.0$, and since most of our detections are associated with galaxies at $z = 1 - 2$, we find them a representative sample for our results.

Following the optical selection criteria of BzK galaxies from Daddi et al. (2010), we extracted a set of BzK galaxies from the Shark catalogue and calculated their median CO SLED together with their 16th and 84th percentiles. The resulting SLED is presented in Figure 4.8. For comparison, we plotted alongside the observational results from Daddi et al. (2015). The Shark simulated SLEDs are in good agreement with the observational points. Both SLEDs peak at the $J_{\rm up} = 4$, observed points lays within the envelope of observed values. The SLEDs predicted form Shark agrees well with the observed SLEDs

of BzK galaxies.

4.4.2 Testing Shark redshift probability calculator with previous detection

Classification of the ALMACAL-CO pilot survey detections relies on the properties of the population generated in Shark Semi-Analytic Models. Since Shark reproduces well the general properties of galaxies in the Universe like galaxy luminosity functions, sub-millimetre number counts and cosmic star formation rate density (see Lagos et al., 2018, 2019) these predictions are reliable. Before applying the probabilistic approach to ALMACAL-CO candidates we checked the predictions on detections of previous surveys: 23 emission lines from ASPECS (Walter et al., 2016; Aravena et al., 2019) and 13 secondary detections of PHIBSS-2 NOEMA survey (Lenkić et al., 2020). We choose detections with confirmed redshift, through other CO lines or optical counterparts.

For each of the detections, we construct Shark-based redshift probability function considering all CO transitions. We compare these predictions with the classification of the sources based on the detection of an optical counterpart or secondary lines and assign them to three categories: observed transition and most probable one agree, the observed transition is one J lower/higher than the most probable one, the difference between observed and predicted is larger than two J levels.

In the case of ASPECS detections, the prediction of the highest probability transition agrees well with the actual transition for 45 per cent of the candidates. However, for most of the remaining candidates, the actual transition was the second or third most probable one. These statistics are more favourable for the PHIBSS-2 detections, where 60 per cent of candidates had the highest probability transition identical to the confirmed one. We stress that in our study we do not choose a single transition with the highest probability, but include all predictions in a probabilistic approach. We summarised the results of the test for both surveys in Figure 4.9.

In general, we observe that the ASPECS and PHIBSS-2 sources are brighter than the sources observed in ALMACAL-CO. Their probability function predicted from Shark is also steeper than the ones assigned to ALMACAL-CO sources (80–90 per cent to almost possible transitions while for ALMACAL-CO sources the maximal is $\sim$ 65 per cent).

The calculator operates on single objects and does not include the parameters of the surveys like frequency coverage or probed volume. With the volume probed by ASPECS in CO (1−0) it would not be possible to observe as high a fraction of low redshift CO(1−0) detections as the redshift calculator suggests. It is also more probable, from the argument of the population, to detect a moderately bright object nearby (CO(1−0) or CO (2−1) z <0.5) than a very bright higher $J_{\rm up}$ transition at higher z (like CO(3−2), $z \sim 1.5$).

4.5 Identification of CO lines candidates

4.5.1 ALMACAL search for rotation lines associated with CO candidates

Single emission line detections in ALMACAL-CO pilot pose a challenge for line classification. However, the advantage of ALMACAL is the availability of wider frequency coverage so that potentially other CO lines could be searching for in the data not included in this pilot survey. To complement the Shark-based redshift probability calculator results, we looked for additional data to help us confirm and identify the detections. Since the ALMACAL-CO pilot consists of a subset of ALMACAL survey data, we looked into the remaining data available for the calibration fields to look for additional CO lines. For each candidate, we assume a CO transition classification and we checked the frequencies of the corresponding CO lines from CO(1−0) to CO(10−9). We centred the new cubes around the observed frequency of the given CO transitions and consider the data within ± 1000 km s^{-1} around it.

The frequency coverage of ALMACAL data differs significantly between the calibrator fields. For the ALMACAL-CO pilot fields checked, the percentage of additional transitions covered by the data ranges from 7 to 33 per cent between candidates, making the probability to detect additional transitions rather low. The resulting coverage of the additional CO transition is uneven for different candidates and cannot be used to exclude some of the classification possibilities.

To estimate if the r.m.s of the new data was sufficient for the CO lines to be detected, we used the prediction of CO emission lines brightness form Shark. Each detection can be assigned a certain CO $J_{\rm up}$ transition for which we generate corresponding average simulated CO SLED. We then compare the expected brightness of the CO line with the r.m.s of the datacube covering a given transition. We check which of the simulated line brightness exceed the signal-to-noise ratio of 3σ, to classify it as detectable. Among all covered transitions, we should be able to detect 48 per cent of the lines. The results differ in between calibrator fields, from 17 to 62 per cent of the covered transition in the field. The summary of this test is presented on Figure 4.10: covered transitions if detectable are marked in green, or in red if the predicted signal does not exceed 3σ of the cube r.m.s. The probability functions of each candidate are plotted alongside to mark most probable detections.

We searched for additional detections in these additional cubes by visually inspecting the data at the expected line frequency and we run *SoFiA* source finder on the cubes. With both approaches, we do not find any detections corresponding to secondary lines of ALMACAL-CO candidates. However, most of the top-probability lines predicted by Shark are not well covered in the available data, therefore the non-detections are not constraining.

The availability of the broad frequency coverage of the ALMACAL fields is a useful tool for providing constraints for the emission lines detection classification. The test presented here will be extended for the full ALMACAL-CO sample search and can become more constraining.

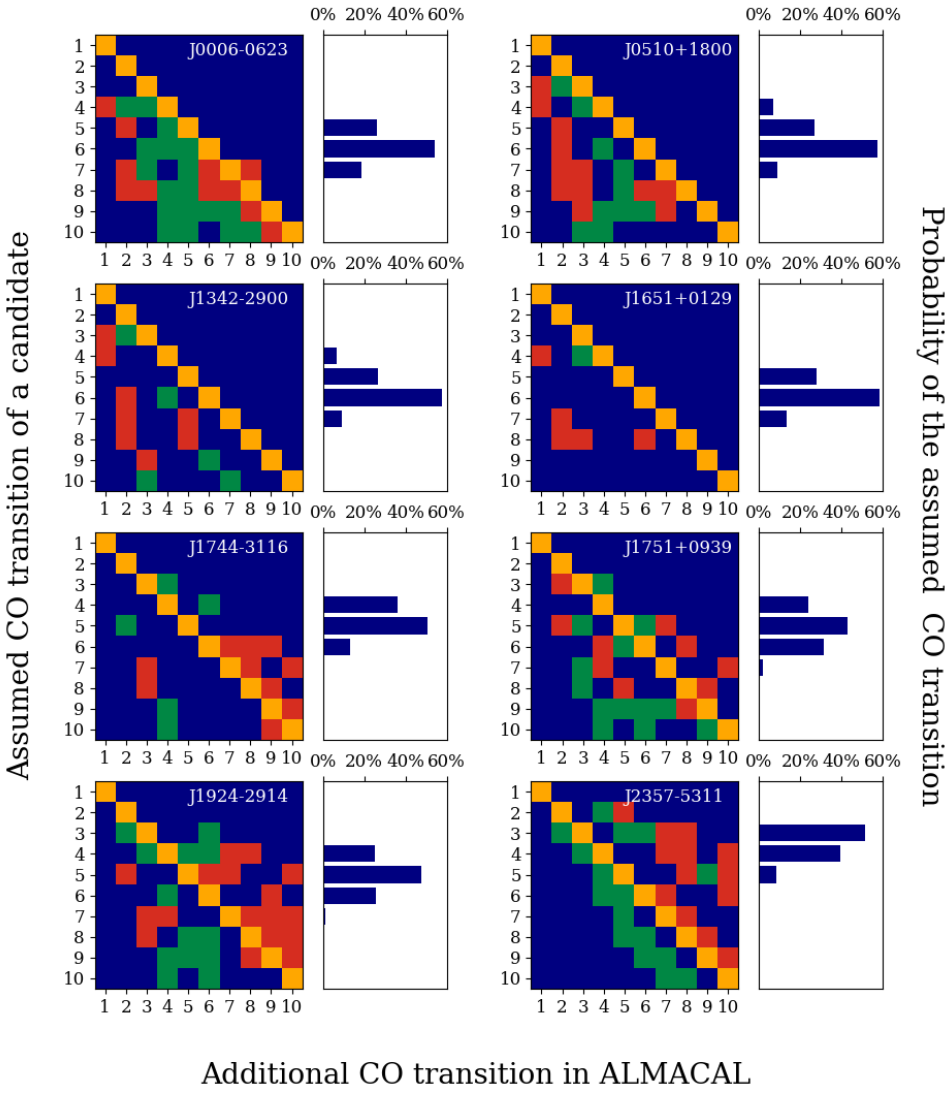

Figure 4.10: The search for additional CO transitions in ALMACAL database. For each source, we check if the ALMACAL data cover the CO ladder corresponding to each probable transition. Based on the r.m.s. of the additional data we check if the predicted transition flux would be detectable wit 3σ significance. Green marks the transitions which flux at the predicted redshift is higher than 3σ of the data noise while the opposite case is marked with red. The non physical transitions are marked in yellow. Finally, the transitions not covered by ALMACAL are marked in dark blue. The panel on the right shows the Shark probability of each transition.

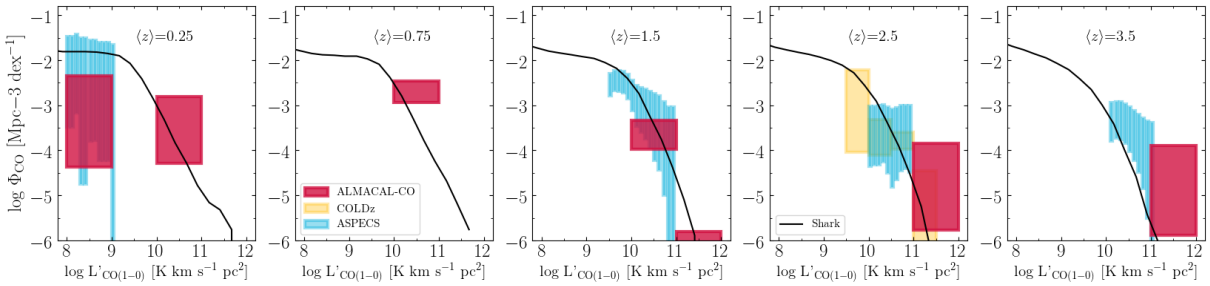

Figure 4.11: The luminosity function based on the ALMACAL-CO detections as the function of CO(1-0) luminosity (red boxes) in different redshift bins. For comparison, the corresponding detections of ASPECS (Decarli et al., 2019) and COLDz (Riechers et al., 2019) are shown in blue and yellow respectively. The black line marks the CO luminosity function of full Shark simulated population (Lagos et al., 2018). We emphasize that the presented pilot survey results are preliminary, and serve as a proof of concept for the further, extended study over the much larger ALMACAL dataset.

4.5.2 Alternative identification of sources as [CII]

With the lack of the optical counterparts to our emission line candidates, we cannot exclude the possibility of the sample contamination by the species different from CO. Five of our detections could be associated with redshift $z \sim 4.5$ [CII] emission. [CII] emission predictions are not included in Shark and are not part of the redshift calculator. As the luminosity function for [CII] at $z \sim 4.5$ is not well constrained observationally (limits by Swinbank et al., 2012; Matsuda et al., 2015) we referred to simulated [CII] luminosity function to estimate the number of expected sources in the volume covered by ALMACAL-CO pilot. Regardless of the model used, redshift $z = 0$ predictions from Herschel by Hemmati et al. (2017), semi-analytical model (Popping et al., 2016; Lagache et al., 2018), the predicted number of [CII] sources in ALMACAL-CO pilot at redshift $z \sim 4.5$ with a luminosity $L = 10^{7.5}$ $L_{\odot}$ is less than one. Additionally, compared to the detection of the ALPINE survey (Bethermin et al., 2020), our detections are much fainter and narrower than any [CII] detections at similar redshift. As the [CII] solution is unlikely, we only include the associations with CO lines in our analysis.

4.5.3 CO(1-0) luminosity function and molecular gas mass density

The ultimate goal of ALMACAL-CO is the measurement of the evolution of molecular gas mass density with redshift and constraining the CO luminosity functions at different redshifts. In this pilot project, we introduce a novel statistical approach relying on the probabilistic redshift association of our sources (see Section 4.3.2). Since every detection is assigned with the redshift probability function, we include all possible CO transitions in the calculations of the CO luminosity function and consequently molecular gas mass density.

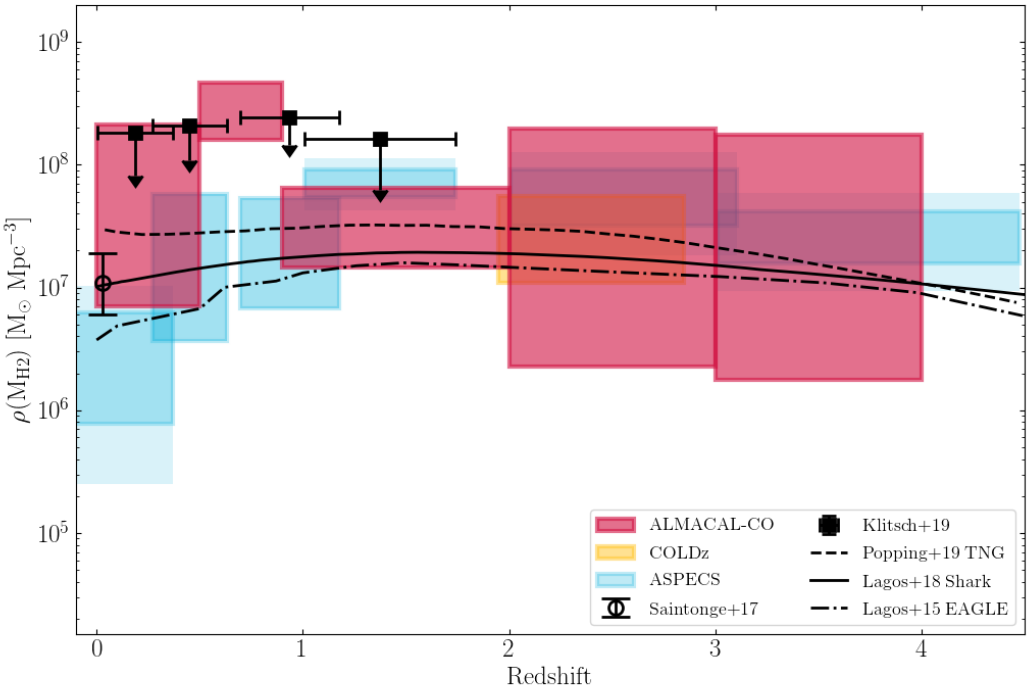

Figure 4.12: The evolution of molecular gas mass density with redshift as measured from the ALMACAL-CO pilot survey (red boxes). The results of ASPECS (Decarli et al., 2019, in blue) and COLDz (Riechers et al., 2019, in yellow) are also shown. Black circle marks the local measurements of molecular mas content of galaxies from xCOLD GASS (Saintonge et al., 2017). Black squares are the limits from molecular absorption search in ALMACAL by Klitsch et al. (2019b). Black lines mark the predictions of evolution of $\rho(H_2)$ from simulations: dashed line - IllustrisTNG (Popping et al., 2019), dashed-dotted line - EAGLE (Lagos et al., 2015), solid line - Shark (Lagos et al., 2018). The simulated evolution represents well the results of ALMACAL-CO pilot, especially at the low redshift bin. The presented results of pilot survey are preliminary and serve as a proof of concept for the search over the extended ALMACAL dataset.

Results of our survey are preliminary, and we are not able to reconstruct the full CO(1−0) LFs with current data. With the future extension to the total ALMACAL dataset, we will improve the statistics and expect significantly more line detections, resulting in better constraints for LFs, and the integrated $\Omega_{\rm H_2}$.

To construct the CO(1−0) LFs for the detections reported here, we calculated 1000 realisations of each of the probability functions, constructing the final statistical sample for our studies. We divided the detections into five redshift bins (resembling the binning of the ASPECS survey): $\langle z_1\rangle$=0.25, $\langle z_2\rangle$=0.75, $\langle z_3\rangle$=1.5, $\langle z_4\rangle$=2.5 and $\langle z_5\rangle$=3.5.

The majority of our detections are classified as high-J transitions. To unify the luminosity function, we used the CO(1−0) brightness predicted by Shark averaged SLEDs corresponding to each detection which we then converted to line luminosity using the assigned redshift and simulated signal integrated flux (Carilli & Walter, 2013).

For each redshift bin, we divide the results into the one-dex luminosity bins. For each luminosity bin, we calculate the luminosity function as:

$$\Phi^{j}_{CO(1-0)}(L') = \frac{1}{V_j}\sum_{i=0}^{N(L')}\frac{R}{c_i}, \tag{4.1}$$

where j correspond to the redshift bin in which we calculate the function, R is the reliability of the detection, assumed for all to be 0.9, i is the iteration over the detections in the luminosity bin and c is the completeness factor evaluated in Section 4.3.4. We repeat the calculations over the all 1000 projections of the probability function of each candidate and eventually, we take the mean value for each luminosity bin. We estimate the errors in the bins following the Poissonian low number statistic uncertainties (Gehrels, 1986).

In the lowest and highest redshift bins, we probe the brightest luminosity objects, brighter then ASPECS and comparable to high redshift COLDz results. We have the best constraints over the bins $\langle z\rangle$=0.75 and $\langle z\rangle$ = 1.5 for which the majority of the points in our sample were assigned to. Especially in the redshift $\langle z\rangle$= 1.5, our results are comparable with ASPECS (see Figure 4.11). We compare the result with the luminosity function predicted by the whole population of Shark galaxies (solid black line on Figure 4.11). Despite sampling the same luminosity function through simulations, we are selecting rare and bright object, which are not dominant in the simulations.

Molecular gas mass is obtained through scaling the CO(1−0) luminosity with the $\alpha_{\rm CO}$ conversion factor. In this work we adopted the Milky Way CO conversion factor $\alpha_{\rm CO}$ = 4.3 $\rm M_\odot/(K\ kms^{-1}\ pc^2)$ for all detections (Bolatto et al., 2013) due to the lack of additional source information. To construct the molecular gas mass function $\rho(\rm H_2)$, each measurement is additionally scaled through the detection reliability parameter and the probed volume.

$$M^{i}_{H_2} = \frac{\alpha_{\rm CO} R}{V_i}\sum_{j=0}^{N}\frac{L'_{\rm CO(1-0)}}{c_j}, \tag{4.2}$$

where the c, R and V are same as in Equation 4.1. We do not integrate over the extrapolated CO Luminosity Functions at any point of the analysis.

Because of the low number of detections, we binned the measurements according to their redshift, following the five redshift bin definition used for luminosity functions. In each redshift bin, we sum the detections' CO(1−0) luminosities scaled with their completeness factor, and calculate the summed molecular mass in redshift bin, dividing by the probed volume, following the Equation 4.2. We repeated that procedure for each of 1000 realisation of the redshift probability function, similarly to the luminosity functions calculations and finally take a mean molecular gas mass density value in each of the redshift bis. Uncertainties for each bin were calculated in the same way as for the luminosity functions.

4.6 Discussion

The ALMACAL-CO pilot project described in this work is a proof-of-concept study, introducing a different approach to the sub-mm untargeted CO emission-line survey. We rely on the data mining of the ALMA archives, focusing on the numerous ALMA calibrators data. With the usage of multiple calibrator fields, we are less prone to the effects of the cosmic variance. However, with this approach, in contrary to the untargeted surveys conducted over the cosmological fields, we cannot rely on the extensive optical ancillary data to identify counterparts and determine the redshift. Anticipating this problem to hold for the full ALMACAL-CO survey, in this pilot we tested a novel statistical approach, based on the probabilistic predictions from Semi-Analytical Models. In this way, we do not focus on the precise classification of each source, instead, we adopt a redshift probability function for each detection and through the statistical evaluation, we arrive at the result averaged over the realisation of the probability function.

The ALMACAL-CO pilot probes different frequencies than previous surveys (ASPECS, COLDz, PHIBSS2), with the majority of the coverage in ALMA Bands 6, 7, 8 (211–500 GHz). By probing such frequencies, we are more likely to observe high-J transitions, which are expected to be fainter than typically found in Band 3 CO(2−1) and CO(3−2) (see Figure 4.4 for probability functions corresponding to each detection).

Thanks to the broad frequency coverage (from Band 3, 84 GHz to band 8, 500 GHz), we can probe several CO lines, from $J_{\rm up} = 1$ to $J_{\rm up} = 6$ (and above, however, due to the low brightness of the high-J transitions we do not expect to be sensitive to those). The ALMACAL-CO pilot survey total volume exceeds that of the previous projects several times. Cumulatively over all fields, the survey is probing a twice larger area of the sky than e.g. ASPECS survey.

Probing the higher-J transitions also means lower expected fluxes of the lines. Our detections are fainter and narrower than the typical sources observed by other surveys (Figure 4.5). This raises a question about the reliability of our detections. The reliability of our candidates was estimated using *SoFiA* intrinsic algorithms, based on the number of negative detections in the inverted data cube. We are aware that we are probing the limits the detection reliability for sub-mm interferometric data and that a fraction of the presented detections is bright noise peaks. At this stage of the project, we are unable to asses the real reliability for these detections, due to the lack of secondary lines and optical

counterparts. We discuss in detail the efforts to understand the nature of our detections in Section 4.5.

The analysis of our results depends strongly on the predictions from Shark Semi-Analytical Models (SAMs). We run a set of tests to check their predictive power. We found that the Spectral Lines Energy Distribution functions (SLED) predicted by Shark, which affect directly our redshift probabilities, are comparable to the ones observed in star-forming galaxies (see Section 4.4.1 for a comparison with Daddi et al. (2015) SLEDS of BzK galaxies). We also run a test on the subset of ASPECS detections with confirmed redshifts (Aravena et al., 2019) and PHIBSS-2 objects with known optical counterparts (Lenkić et al., 2020) through our redshift probability calculator (details of the test can be found in Section 4.4.2). The probability functions assigned to detections of these surveys predict the correct transition in about 50 per cent of the cases. In any case, the actual transition was assigned with the null probability.

The primary goal of the ALMACAL-CO survey is the assessment of the evolution of the molecular gas mass density and determination of the CO luminosity functions at different redshifts. We derived the CO(1−0) luminosity functions (LFs) based on our detections in five increasing redshift bins. In each redshift range we probe bright objects $L > 10^{10.5}$ [K km s^{-1}pc^2]. Additionally, we detect low luminosity CO(1−0) in the $0.0 < z < 0.5$ bin, in agreement with results of ASPECS, results in the same redshift range. Especially in the redshift range $z > 2.0$, we probe the bright end of the distribution, not covered by previous surveys. The LFs derived in this work illustrate the concept of the ALMACAL-CO pilot survey and should not be seen as fully sampled at this stage. Further analysis of the complete data set is expected to provide tight constraints and a good sampling of the CO(1−0) LFs in different redshift bins.

We convert the derived CO(1−0) luminosities, scaling by the volume probed by the survey as well as completeness and reliability coefficients, to molecular gas mass density. As the LFs covered in ALMACAL-CO pilot are not complete, we do not attempt to integrate over them, instead, we directly convert the luminosities into a molecular gas mass as shown in Equation 4.2. The derived molecular gas mass density evolution with redshift is consistent with findings of previous surveys (ASPECS, PHIBBS-2 and COLDz) for the redshift above $z > 1.0$.

The largest difference between the evolution of $\rho(H_2)$ presented in this chapter and findings of previous surveys arises for the redshift below $z < 1$, where the effects of cosmic variance are the strongest and measurement is higher than ASPECS survey but are consistent with the local measurements of Saintonge et al. (2017) and the upper limits from the CO absorption search from (Klitsch et al., 2019b). We reject the hypothesis that these low redshift objects are misclassified $z = 4.5$ [CII] detections, based on the low expected line luminosity compared to the results of the ALPINE survey (Bethermin et al., 2020) at the same redshift (see full discussion in Section 4.5.2).

The evolution of $\rho(H_2)$ predicted by the hydrodynamical simulations like EAGLE and IllustrisTNG (Lagos et al., 2015; Popping et al., 2019) is lower (especially for $z > 1$) than any of the observations so far, including our results. Underpredictions of the molecular gas mass is a known issue of the hydrodynamical simulations, which predict less dramatic

evolution of molecular gas mass density than observed so far. Predictions of $\rho(H_2)$ shape vary between the simulations, especially in the low-z regime, predicting a sharper or shallower decline of the $\rho(H_2)$ towards lower redshift. Altogether any of the simulations are not reproducing observed sharp evolution of the molecular gas mass function observed by ASPECS. Our results are more in agreement with previous observations and are above the predictions of the simulations. In the analysis we use the simulated CO SLEDs from the Shark SAMS, but our results do not depend on the Luminocity Function or molecular gass mass density predicted by these simulations.

For comparison, we have also plotted the $\rho(H_2)$ evolution predicted by Shark SAMs. The molecular gas content of galaxies in the simulation is calculated independently on the CO luminosities and is not the result of the LFs integrations. Shark is also anchored to the $z = 0$ measurements (Saintonge et al., 2017), but similarly to other simulations, the predicted evolution of gas mass density is less dramatic then observed. Overall our results agree more with the previous observational findings then the simulated evolution of cosmic molecular gas mass density.

The results of ALMACAL-CO taken together with other surveys are still broadly consistent with two different scenarios: either the evolution of molecular gas with redshift follows the evolution of SFH or there is no evolution, similarly as for $\Omega_{\rm HI}$. Within the errors, our results are consistent with the previous observations but are higher than simulation predictions. The detailed shape of the molecular gas mass curve requires additional observations to understand if the function follows the SFH of the Universe or is flat similar to the $\Omega_{\rm HI}$).

The pilot project presented in this work was primarily conducted to demonstrate the feasibility of using ALMACAL calibrator data for untargeted emission-line surveys. As calibrator data are not as homogenous as the science program observations, the noise properties and uneven frequency coverage make the survey more challenging in some aspects. Nevertheless, in the subset of the deepest observations, we detected eleven candidates making the case for an ALMACAL-wide search. After the pilot, we will expand the untargeted search to all 880 calibrator fields including all new data collected during the past two years (since the original selection of pilot sample). This will expand the coverage of the survey to ~ 600 arcmin^2 and increase the volume ~ 50 times. With this significant expansion of the survey, we expect to detect 50–100 CO emitters, improving the statistics significantly.

4.7 Conclusions

We have presented the results of an ALMACAL-CO untargeted pilot CO emission-line survey, using the ALMA calibration data from the ALMACAL project. We report eleven line detections found in the 38 calibrator fields that each have an accumulation integration time exceeding 40 minutes. A characteristic of our survey is that our fields are centred on bright sub-mm ALMA calibrators and are not cosmological fields that have been observed in a wide range of wavelengths and hence we lack ancillary data to classify the detections. Therefore we adopt a probabilistic approach by constructing a redshift probability func-

tion for each of the candidates. Based on the predictions from the Shark Semi-Analytical Models, we assign corresponding CO transition probabilities to each of the detections. Although our results are dependent on SAMs we show that the final results are not strongly dependent on the parameters of the simulations and are in line with results of previous observational surveys. Because our survey covered more of the high-frequency bands (Bands 6, 7, 8 covering frequencies between 211 and 500 GHz), we improved on the total volume coverage of previous surveys and have significant volume coverage for the high-J transitions ($J_{\rm up} = 5, 6$).

The CO luminosity functions and the cosmic molecular gas mass density derived from our ALMACAL-CO survey follow the trends indicated by the previous untargeted surveys. We report most of our detections to be more luminous than predictions from simulated CO luminosity functions. Our molecular gas mass function is consistent with the results of other untargeted surveys. ALMACAL-CO pilot is a proof-of-concept project and the results presented in this work are preliminary.

In the future, we are aiming to confirm the ALMACAL-CO pilot detections and constrain their redshifts by searching for optical/IR counterpart as well as by conducting targeted ALMA follow-up for other CO emission lines. After the success of the pilot study, we are expanding the untargeted survey to all ALMA calibrators (over 880 fields). To complement the process of collecting the ancillary data of the CO-selected galaxies we will use the Semi-Analytical Models (SAMs) to place constraints on the detections identification and use them as predictions for the follow-up observations.

Chapter 5

Conclusions and Future Prospects

5.1 Thesis summary

Galaxies are not systems evolving in isolation, on the contrary, they are actively interacting with the surrounding medium. The flows of gas are an important aspect of the evolution of galaxies, star formation processes and quenching transformation. The diffuse gas accretes onto galaxies through the Circumgalactic Medium (CGM), replenishes the reservoirs of gas, and further converts into a molecular gas, is the actual fuel of Star Formation. Describing these processes, characterizing the gas content of galaxies' ISM and CGM is vital for the understanding of the evolution of galaxies and their star formation processes from the early Universe until today. In this thesis, I presented two projects investigating a cycle of baryons at two scales and their impact on the evolution of galaxies: the local baryon cycle in diffuse CGM halos, traced by the quasar absorption systems, and the global baryon cycle related to the molecular content of galaxies across the cosmic time in an untargeted CO emission line survey.

To understand how galaxies replenish their gas, vital to sustain the star formation, we study the local baryon cycle - the gas flows through the CGM. Due to the low surface brightness, direct studies of the CGM's diffuse gas are challenging. Nevertheless, we can probe these vast gaseous halos around galaxies using bright distant point sources such as quasars, by identifying the hydrogen and metal absorption lines in their spectra and connecting them to the host galaxies detected in emission. By combining the emission and absorption tracers we can characterize the properties of galaxies associated with the quasar absorbers and study their interactions with the gas in CGM.

To address the problem of identification and characterisation of absorber host galaxies we have built the multi-wavelength MUSE-ALMA Halos survey (Chapter 2). The survey is a collection of MUSE observations of five quasar fields, supplemented with archival quasar high-resolution UV spectroscopy from VLT/UVES and Keck/HIRES, for absorber identification. I find that most absorption systems can be associated to pairs or even groups of galaxies, which members are both star-forming and quiescent galaxies. Combining the emission and absorption tracers we can characterize the properties of galaxies associated

with the quasar absorbers and study the gas in their CGM.

The gas travelling through the CGM enters a galaxy and replenishes its gas reservoirs, which further transforms into the molecular phase - the direct fuel of star formation. We know that star formation takes place in dense molecular clouds, however, the link between the general content of the molecular gas of galaxies with the global star formation of the Universe is complex.On the global scales, the star formation rate history (SFH) of the Universe changes with time and it is not clear what drives the $z \sim 2$ peak and subsequent decline. Recent studies suggested a possible link between the SFH and the cosmic density of H_2 - the most abundant molecule in the Universe.

The H_2 molecule cannot be directly traced in the extragalactic sources due to its very weak vibrational transitions, which additionally cannot be excited in the low-temperature environment of Giant Molecular Clouds. Therefore, to quantify the molecular gas content of galaxies, we use a molecular gas tracer, the CO molecule with several bright transitions. By observing the CO emission lines and converting their brightness to the luminosity of the lowest possible CO transition, CO(1−0), we obtain the molecular gas mass through scaling the CO luminosity with the CO-to-H_2 conversion factor $\alpha_{\rm CO}$. Today, the uncertainty of that conversion factor is one of the major unsolved problems in sub-millimetre extragalactic astrophysics.

Measurements of the molecular gas content of galaxies were so far mostly conducted through targeted surveys of a large sample of star-forming galaxies. Such studies have a high detection rate, but probe only the most massive objects, and do not account for the whole population of the CO-bright or molecular gas-rich galaxies. To complement existing deep pencil-beam surveys I employed an alternative method and embarked on an archival data survey, utilizing ALMA calibration data from the ALMACAL project to construct an untargeted CO emission line ALMACAL-CO survey. ALMACAL is aiming to use numerous ALMA calibrations as scientific data (`almacal.wordpress.com`). Since every ALMA science observation is accompanied by several calibration measurements, ALMACAL provides a vast and diverse dataset suitable for multiple science cases. ALMA calibrators are usually bright sub-millimetre quasars (mostly blazars) and are straightforward to extract from the data, resulting in a science image of the calibrator field, which for many science cases provide a random sampling of the Universe. Thanks to the uniqueness of the ALMACAL dataset we can study galaxies over a wide area and are less sensitive to the effects of the cosmic variance than the pencil-beam surveys.

In ALMACAL-CO pilot (Chapter 3 and Chapter 4), we chose from amongst all ALMACAL fields, the ones with the longest total integration time (38 fields, observation time over 40 min each) and detected eleven emission lines. Because for our sample no large-scale auxiliary data are available and we lack the optical counterparts or secondary emission lines, we cannot simply assign redshift to our detections. To complement for the lack of information I implemented a probabilistic approach to assign CO transitions to the detections, based on simulated galaxy population from the Shark Semi-Analytical Models (SAMSs). Using the predicted CO luminosities of the simulated galaxies we built a CO transition probability calculator, assigning a transition probability function to each of the candidates by comparing the candidates' flux with the one expected for the final calcula-

tion of the CO transitions in simulations. Eventually, each candidate was assigned with a range of probabilities, all of which were included in the calculations of the CO luminosity function and the molecular gas mass.

Based on the results of that probabilistic approach and including the cosmic volume probed by the ALMACAL-CO pilot survey I calculated the molecular gas mass density evolution with redshift, represented by the candidates. The evolution of the molecular gas mass density derived from the ALMACAL-CO pilot is in agreement with the constraints put by the previous untargeted surveys (e.g. ASPECS, COLDz). At the same time, it is also consistent with the lack of any evolution of molecular gas mass density.

The pilot survey aimed to test the capability of the ALMACAL dataset for its suitability to be used for an untargeted emission line survey. Although the measurements are not yet constraining I prepared the methodology alongside with the probabilistic approach complementing the identification of the optical counterpart, which can be expanded to the whole ALMACAL database. I expect ten times more detections allowing to place tighter constraints on the evolution of the molecular gas mass density of the Universe.

From studies of the gas flows and the distribution of metals in the CGM to the molecular gas content of the galaxies, in my work, I trace the cycle of baryons on local and global scales. The local cycle, of the CGM of galaxies allows for studies of the flows of matter between the IGM and the CGM as well as the metallicity of the matter, an important aspect of the evolution of galaxies. Finally, the evolution of the molecular gas density follows the relation between baryons and the star formation history of the Universe. Thought the surveys described in this work I try to answer the questions posed in the Introduction: How galaxies accrete the gas, and how the metals are distributed througout ISM and CGM, what drives the star formation history of the Universe and how the molecular gas content of galaxies changes with redshift. All the above concepts summarise the cycle of baryons in the Universe across time.

5.2 Future work

5.2.1 MUSE-ALMA Halos

The MUSE-ALMA Halos project presented in this thesis (Chapter 2) shows the power of the statistical approach to characterise of the host galaxies of quasar absorbers (Hamanowicz et al., 2020). Despite a large number of known hydrogen and metal absorbers (with a range of probed column densities), only a handful of them have well-studied associated galaxies. The study so far focused on particular types of absorbers, such as DLAs (e.g. Christensen et al., 2007; Fynbo et al., 2010; Augustin et al., 2018), strong MgII absorbers (e.g. Nielsen et al., 2018; Zabl et al., 2019) or metal-rich systems detected in molecular gas emission (e.g. Neeleman et al., 2016, 2018; Fynbo et al., 2018). To describe in detail the properties of gas flows in the CGM, trace the distribution of metals, disentangle inflows and outflows and finally understand the connection between the gas and galaxies we need an unbiased statistical sample of absorbers and host galaxies, irrespective of their properties.

Studies with IFU spectrographs provide an unprecedented way to detect the galaxies associated with absorbers in emission as well through the continuum detections and provide a spectroscopic and kinematic information about the objects. Recent studies (Klitsch et al., 2018; Péroux et al., 2019) have shown the potential of combining multiwavelength observations of the absorber systems and their host galaxies. The star formation rates, metallicities and kinematics of ionised gas can be obtained from the IFU spectrographs (VLT/MUSE). The molecular gas content, its kinematics and depletion times come from sub-millimetre observations (ALMA). All above are complemented by the high spatial resolution multi-band imaging providing the detailed information about the morphologies and stellar content of these galaxies through the Hubble Space Telescope (HST) observations. The best high spatial resolution observations are achieved by the Hubble Space Telescope. Our group has granted time for an extensive multi-band observational campaign with the HST/WFC3 camera, covering ~ 40 fields of the $z < 0.85$ absorbers, covering the expanded MUSE-ALMA Halos sample observed also by MUSE. These observations will allow us to address following aspects of the CGM studies.

We are using a multiwavelength approach, combining the techniques mentioned above in the MUSE-ALMA Halos extension to create the first multiwavelength statistical sample of the host galaxies of quasar absorbers at $z < 1$ (Rahmani et al., 2018a,b).

The first step is the expansion of the MUSE-ALMA Halos sample to the new fields of quasar absorbers observed with MUSE. A completed multi-period observation project provided MUSE observations of additional ~ 40 absorbers. We will implement the analysis described in Chapter 2, and complement it with the 3D kinematical modelling (Bouché et al., 2015; Péroux et al., 2016) to build the full picture of the properties of ionized gas in the CGM of these galaxies. Alongside these measurements we obtain the lacking UV quasar spectra from VLT/UVES for the quasars missing high-resolution UV spectroscopy, to identify new absorbers, measure the metallicity of the gas and disentangle its multi-component kinematics, to further compare it with the results of the ionised gas kinematic modelling.

In past years with newly avaliable IFUs many HI-rich absorbers were found to have multiple associated galaxies contrary of the traditional view of a single galaxy-absorber pair (e.g. Bielby et al., 2017; Klitsch et al., 2018; Péroux et al., 2019; Hamanowicz et al., 2020). These discoveries point toward a possible origin of the gas detected in absorption as remains of tidal interactions of the galaxy group or as the gas filling the space between galaxies aligned in a filamentary structure. The origin of that gas can be determined through the detailed morphological studies of these galaxies.

Finally, to complete census of baryons in the hosts of quasar absorbers we complement the studies with the measurement of the molecular gas emission with ALMA, which will provide molecular gas masses. So far, a subset of systems described in Chapter 2 were observed with ALMA, targeting the CO(2−1) and CO(3−2) emission lines (Augustin et al., 2018; Péroux et al., 2019, Szakacs et al. in prep). ALMA observations provide also kinematic measurements of molecular gas allowing for the comparison of the dynamics of the ionised and molecular gas in search of the signatures of gas flows in the HI-rich galaxies. Together with the measurements of stellar and dynamical masses, the molecular

gas mass leads to the measurements of the gas fraction of these galaxies (Fynbo et al., 2018) and measurements of the depletion time of the molecular gas, tracing the star formation processes in these galaxies (Tacconi et al., 2020).

Combining the multiwavelength view of the host galaxies of quasars absorbers we will complete the baryon census of these systems, study the group/filament origin of the gas detected in absorption and study the link between the host galaxies and flows of matter through the CGM.

5.2.2 The ALMACAL-CO untargeted CO emission line survey

The ALMACAL-CO pilot survey described in Chapter 4 shows the potential of archival calibration data usage as an efficient untargeted emission line survey. Using only a subset of fields and frequency coverage we detected eleven candidates, two with optical counterparts. This exercise advocates for the exploration of the ALMACAL-wide survey, with an expanded frequency coverage and number of fields. Since the original data selection for the pilot, the database expanded with two yearly cycles of ALMA observations, potentially extending the integration time over the pilot fields.

The full ALMACAL-CO survey will include 880 calibrator fields, expand the frequency coverage and, consequently the probed volume. Extrapolating from the number of sources detected so far we can expect to detect approximately 100 sources. Although most of the fields will not be deeper than the ones observed in the pilot, we will include a much wider frequency coverage, expanding the volume 50 times. The numerous calibrator fields distributed over the ALMA sky additionally provides further robustness against the effects of cosmic variance, especially at low redshift. Furthermore, the expanded frequency coverage of ALMACAL-CO will include the search of [CII] emission from galaxies above $z > 4$, potentially leading to the first observational constraints on the [CII]-emitter luminosity function.

One of the main challenges of the ALMACAL-CO survey is the redshift determination of the observed emission lines. In the next step, we will confirm the ALMACAL-CO pilot detections and constrain their redshifts by searching for optical/IR counterparts as well as through targeted ALMA follow-up for other CO emission lines. To complement the long term process of collecting the ancillary data of the CO-selected galaxies we will use the Semi-Analytical Models to place constraints on the detections identification and use them as predictions for the follow-up observations.

5.3 Final remarks

The evolution of cosmic baryons and its relation to the star formation history of the Universe is a central point of the global view of the galaxy evolution. With future facilities, studies of the cosmic baryon cycle will enter a new era. With the SKA (Square Kilometer Array, Dewdney et al., 2009) radio interferometer, we will map the HI 21-cm emission in galaxies and groups environment up to higher redshifts than what is achievable to-

day ($z > 0.2$). MUSE and future generation of IFU spectrographs (e.g. HARMONI) on the Extremely Large Telescope (Theuns & Srianand, 2006; Thatte et al., 2010; Augustin et al., 2019), will trace faint emission of the CGM around typical star-forming galaxies and possibly for the first time enable capturing the clumpy nature of the CGM in emission. Furthermore, the upcoming public surveys with 4MOST (de Jong, 2019), a multi-object spectrograph designed for the VISTA telescope, will provide a SDSS-like wide-field optical survey of the southern sky, resulting in first deep spectroscopical catalogues of high-resolution quasar spectra and as many as 800,000 spectra of galaxies up to redshift $z \sim 1$ (WAVES survey, Driver et al., 2019). Such a vast database of spectroscopic information for both quasars and galaxies will allow for the cross-correlation of the quasar absorbers and their host galaxies providing an unprecedented number of absorption systems incomparable to any sample we obtain today. Such studies of the gas kinematics and metallicity and the general CGM properties of the absorber host galaxies allow for better constraints of galaxy evolution models. Additionally, the new generation X-ray telescope, Athena (Barcons et al., 2012), will provide a high energy view of the WHIM and CGM and the links between the hot IGM and the cool CGM gas around galaxies and groups.

Multi-wavelength observations of CO-bright galaxies and comparison with simulated galaxy populations will provide valuable insight into the processes driving the star formation and regulation of the molecular gas content of galaxies in the Universe. Characterisation of the CO-bright galaxies will also help to understand the distant ($z > 6$) sources, which are detected with ALMA today and will be discovered by the JWST in the future. Combining the observations with the hydrodynamical simulations we will also address the question of the non-uniqueness of $\alpha_{\rm CO}$ conversion factor and its evolution with redshift. Additionally, high-resolution studies with ALMA of galaxies at the peak of the star formation history ($z \sim 2$) will lead to an understanding of the star formation processes in galaxies across the cosmic time.

Altogether, the future telescopes and space missions will bring us closer to understanding the star formation processes in the Universe and the impact of cosmic baryon cycle on the evolution of galaxies.

Appendix A

Galaxy properties

We present HST WFPC2 images with positions of galaxies associated with the absorbers described in this work (Figures A1–A6), as well as their spectra and continuum detection on MUSE white light images (A7–A14). MUSE white light and HST images for each quasar field present all detected galaxies associated with absorbers in our sample, differentiated by colours. Both images are in the same scale frame and oriented North up, East to the left. For clarity, we show the Q1130z031 rich group members on a separate figure (for details on this system see Péroux et al., 2019).

Each spectrum panel gives the plot of the MUSE spectra of the galaxies from 4800 Å to 9300 Å except for the already published MUSE studies of Q0152z038 (Rahmani et al., 2018a), Q0152z078 (Rahmani et al., 2018b), Q1130z031 (Péroux et al., 2019) and Q2131z043 (Péroux et al., 2017). Emission lines in the galaxy are marked in red, as are the absorption lines of interstellar CaII H&K from the galaxy. The spectrum of sky lines before substraction is plotted in red. In the upper left-hand corner of each galaxy spectrum frame appears the identifier of the galaxy (first four digits of QSO right ascension, two decimal points of the absorber redshift (e.g. z019 for $z = 0.19$), and the galaxy ID (e.g. G1) with which the galaxy spectrum is associated; and the impact parameter (in kpc) of the QSO field images (Figures A1 to A6) projected onto the plane perpendicular to the sightline of the QSO. Two postage stamps of the galaxy appear to the right of this label, above the spectra (from MUSE and from the WFPC on HST). If the HST inset is not present, the particular galaxy is outside the field of view of HST/WFPC2 camera. To the right of the galaxy spectrum is the spectrum of MgII λ2796 or FeII λ2600 for the absorption system noted in the galaxy spectrum frame (in rest-frame velocity units). The blue line is the normalized UVES/HIRES quasar spectrum, red is a Voigt profile fit to the line. Because of the velocity span of some of the galaxies, not all panels are presented on the same velocity scale. We do not present among the spectrum figures the spectra of the galaxies extracted from the quasar PSF (Q1211z062-G1 and Q1232z039-G1), because we were only able to retrieve the [OII] emission line with false continuum; they are presented in Figure 2.2.

In the QSO fields (Figures A1 to A6), each galaxy associated with an absorber is marked with a circle, colour-coded by the absorber ID. The name of each absorber and a matching

coloured dot matching appears in the upper-left-hand or lower-right-hand corner of the images of the QSO fields.

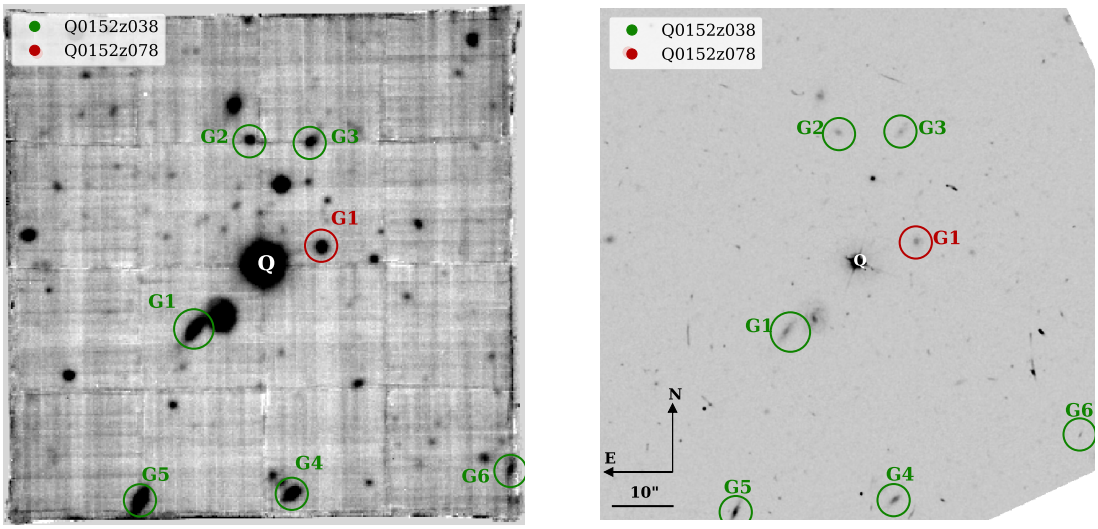

Figure A.1: MUSE white light (left) and HST F702W (right) images of Q0152-2001 quasar field with galaxies associated to all absorbers marked. Green circles mark galaxies associated with Q0152z038 absorber, red - Q0152z078 absorber. Q indicates the quasar.

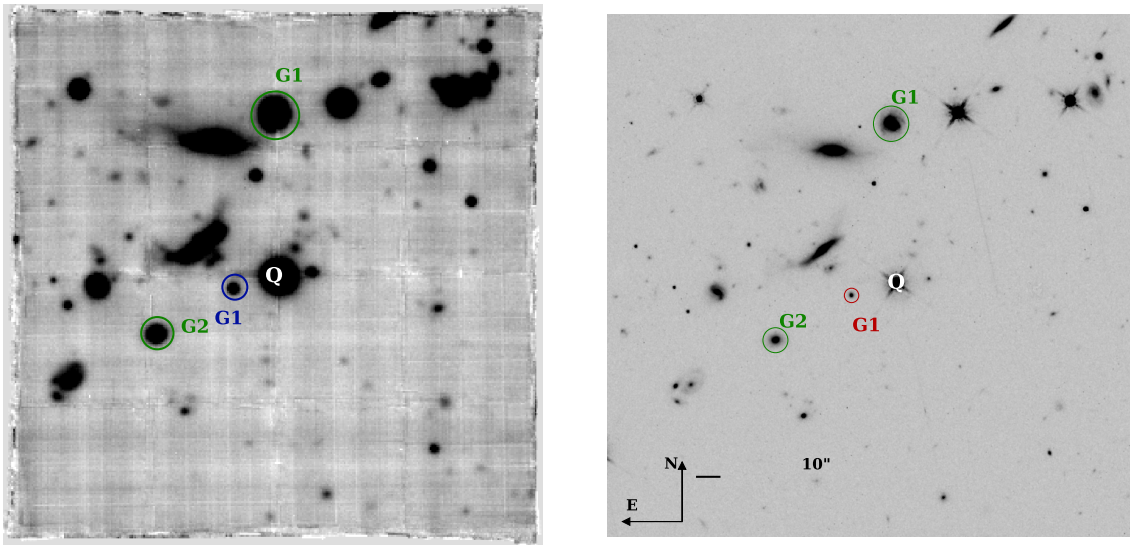

Figure A.2: MUSE white light (left) and HST F104W IR (right) images of Q1130-1449 quasar field with galaxies associated to two absorbers marked. The blue circle marks the galaxy associated with Q1130z019 absorber, green circle with Q1130z032 absorber. Q indicates the quasar. The third absorber in this QSO, Q1130z031, is shown in Figure A.3

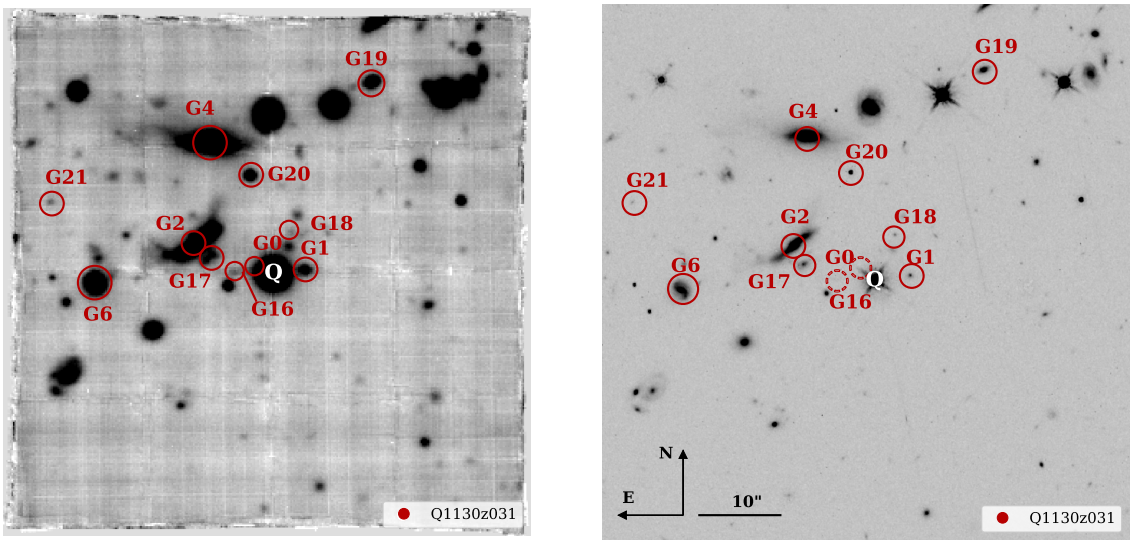

Figure A.3: MUSE white light (left) and HST F104W IR (right) images of Q1130-1449 quasar field with galaxies associated to Q1130z031. Positions of galaxies not detected on the HST image (G0, G16) are marked with dashed circles. Q indicates the quasar.

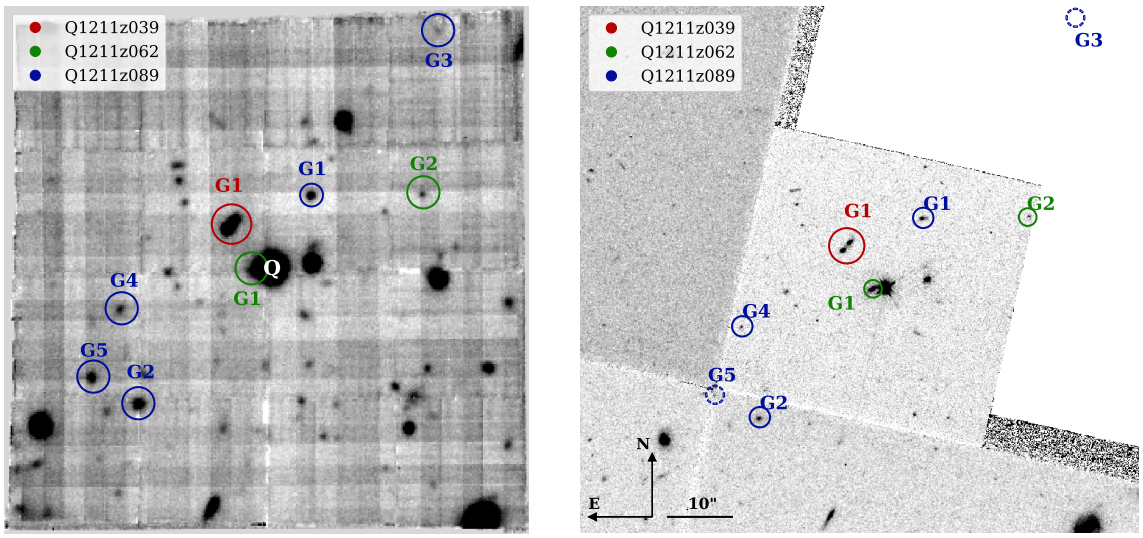

Figure A.4: MUSE white light (left) and HST F702W (right) images of Q1211+1030 quasar field with galaxies associated to all absorbers marked. Red circles mark galaxies associated with Q1211z039 absorber, green - Q1211z062 absorber, blue - Q1211z089 absorber. Q indicates the position of the quasar.

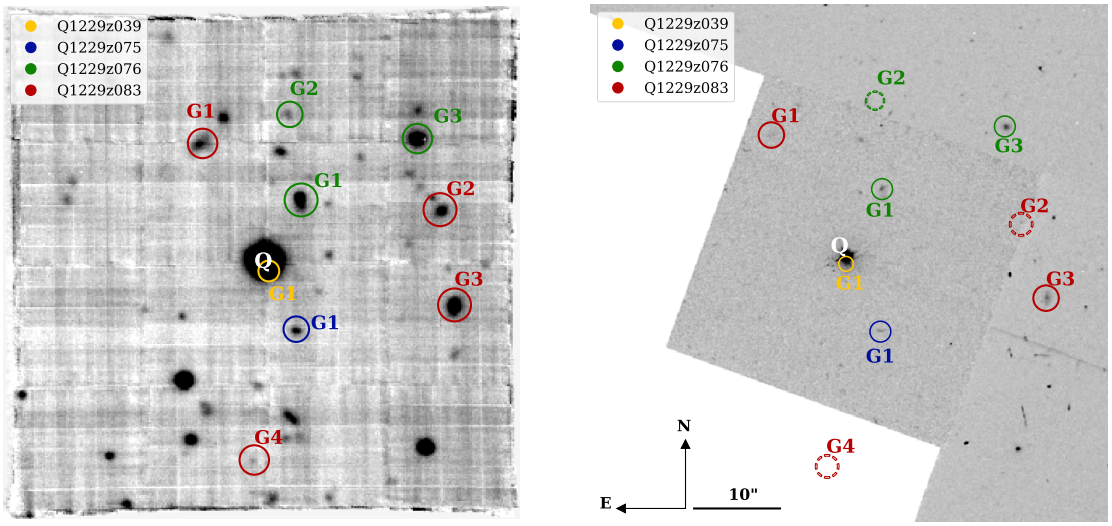

Figure A.5: MUSE white light (left) and HST F702W (right) images of Q1232-0224 quasar field with galaxies associated to all absorbers marked. Yellow circles mark galaxies associated with Q1232z039 absorber, blue - Q1232z075 absorber, green - Q1229z076 absorber and red with Q1232z083 absorber. Q indicates the quasar.

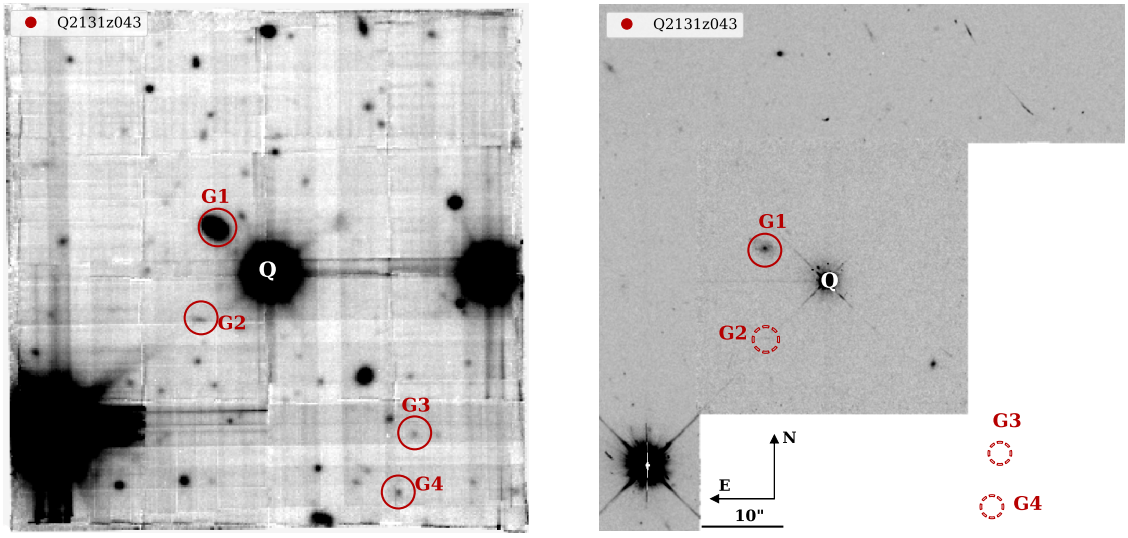

Figure A.6: MUSE white light (left) and HST F702W (right) images of Q2131-1207 quasar field with galaxies associated to an absorbers marked in red. Dashed circles mark the non-detections. Q indicates the quasar.

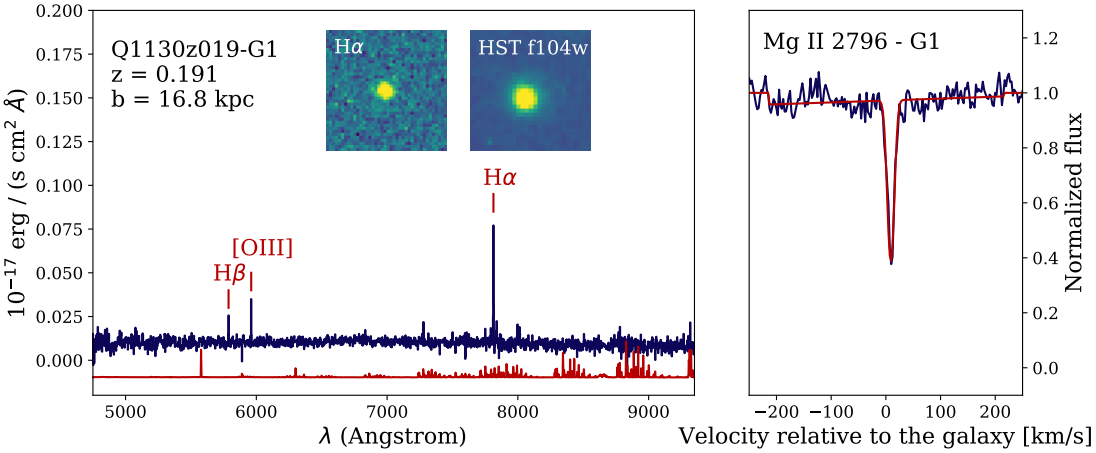

Figure A.7: Q1130z019 absorber and associated galaxy (continuum images A.2).

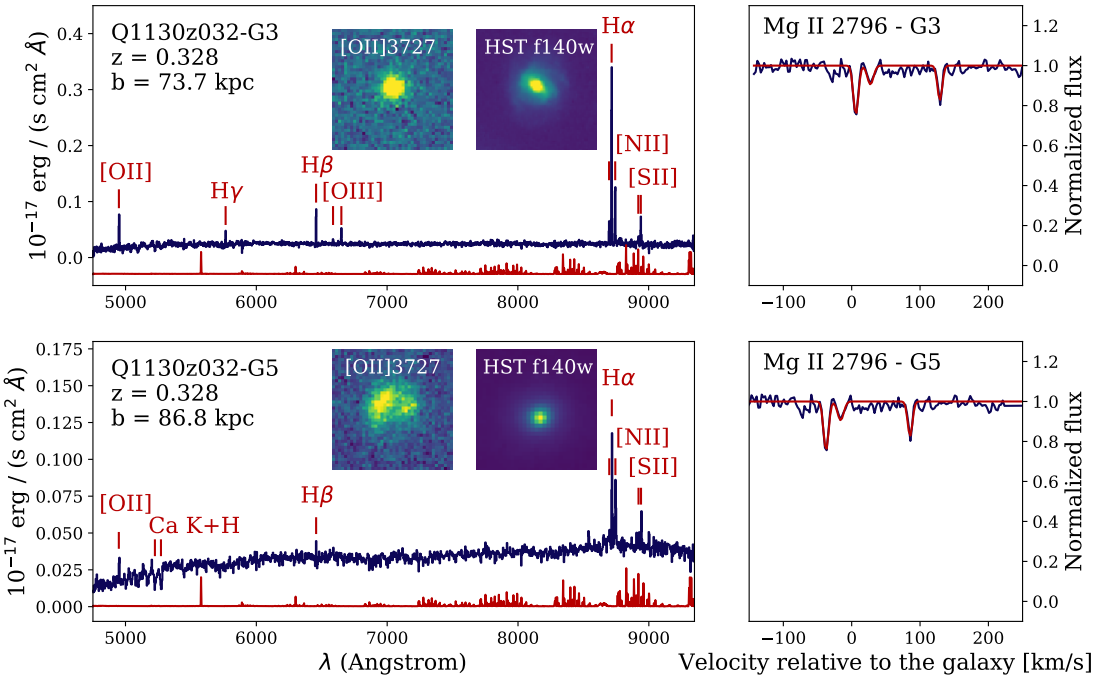

Figure A.8: Q1130z032 absorber and associated galaxies continuum images A.2).

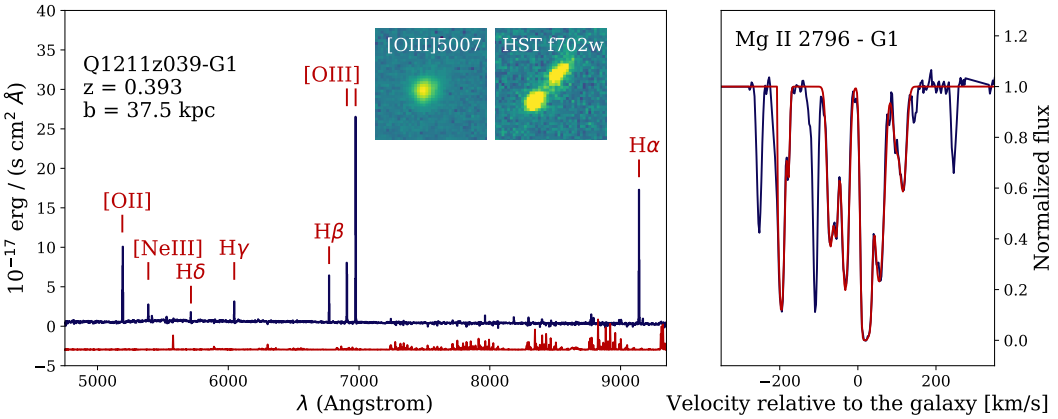

Figure A.9: Q1211z039 absorber and associated galaxy (continuum images A.4).

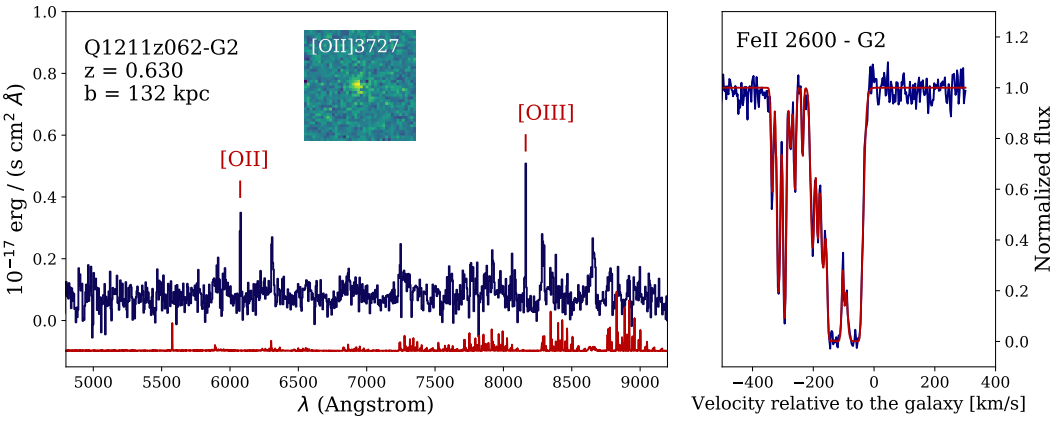

Figure A.10: Q1211z062 absorber; G1 is a low impact parameter associated galaxy, and only [OII] emission line was extracted form the quasar PSF (presented on the Fig. 2.2. Here we present only the spectrum of the second associated galaxy (G2). For this absorber MgII λ2796 falls in the UVES spectral gap, we plot FeII λ2600 profile instead (continuum images A.4).

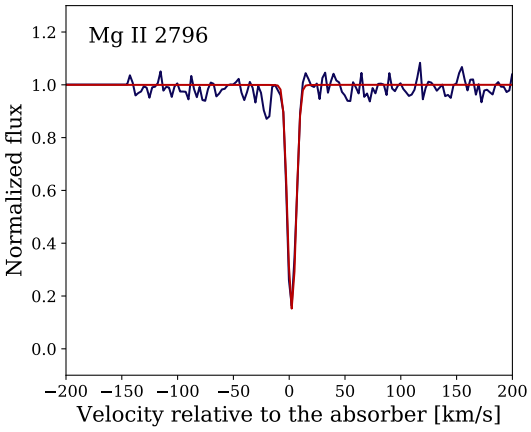

Figure A.11: Q1211z105 absorber MgII profile. This is the only absorber in the sample for which no associated galaxy was found.

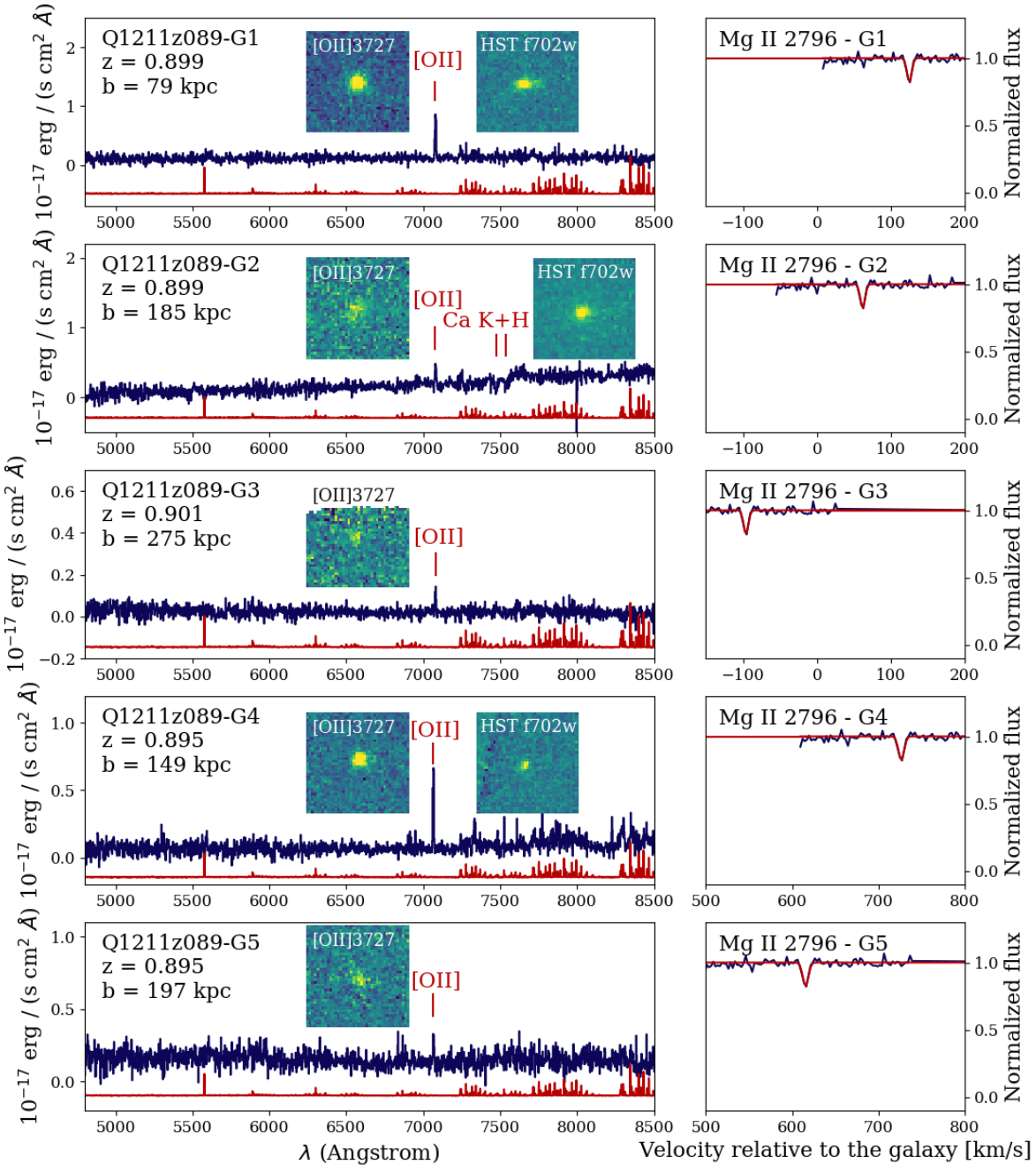

Figure A.12: Q1211z089 absorber and associated galaxies. G3 is outside HST/WFPC2 field of view, G5 lays on the edge of the WF and PC chips of the HST WFPC2 camera and is not detected in continuum (continuum images A.4). Note that due to larger velocity offset, G4 and G5 are presented in different velocity range than G1-G3.

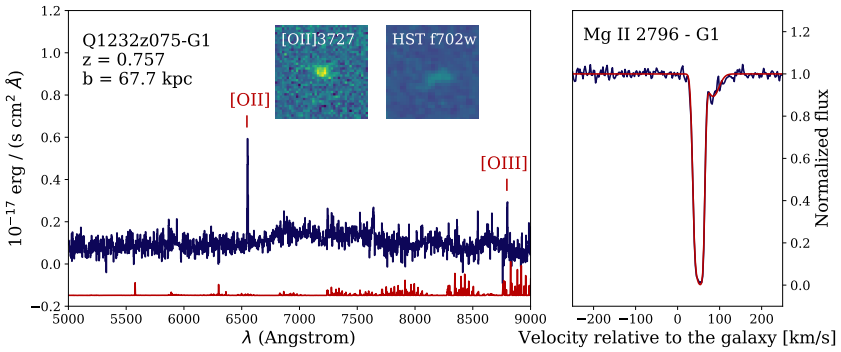

Figure A.13: Q1232z075 absorber and associated galaxy (continuum images A.5).

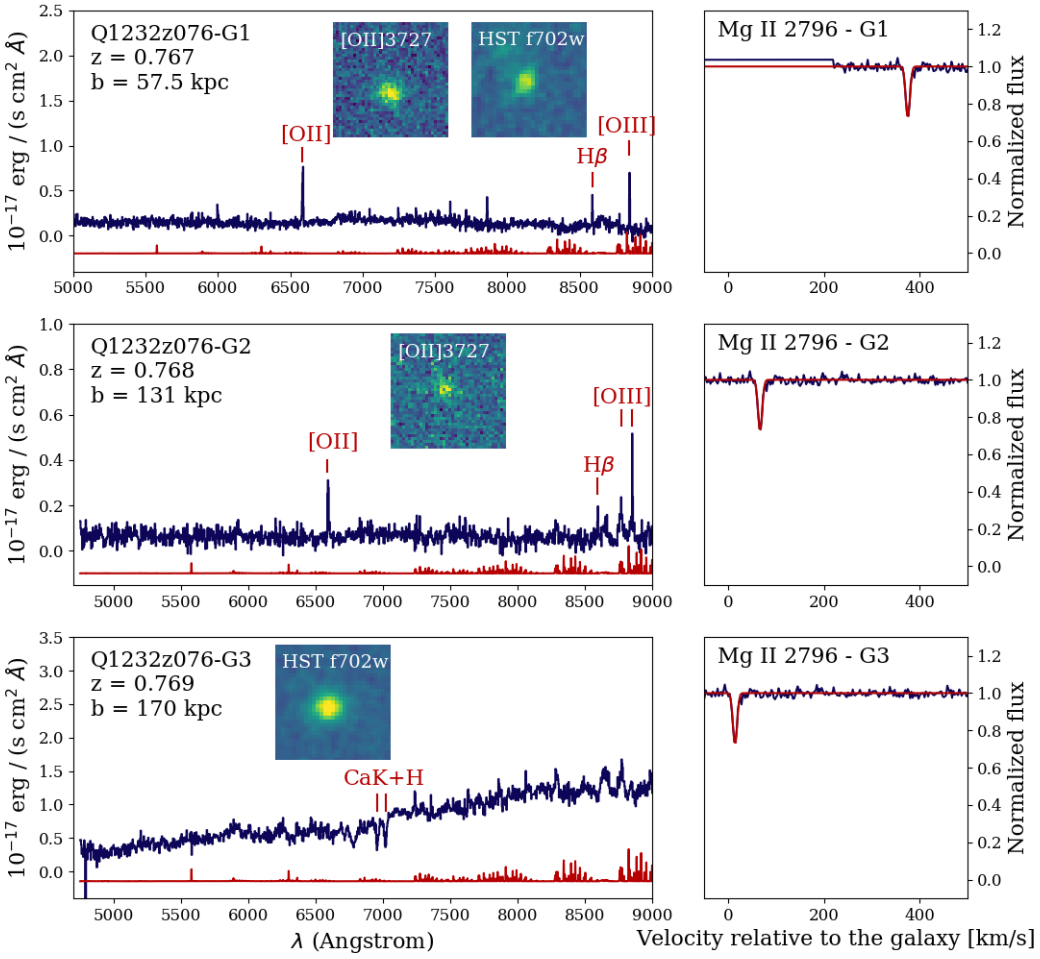

Figure A.14: Q1232z076 absorber and associated galaxies. G3 is a quiescent galaxy with only absorption line and was detected only in continuum in MUSE whit light image (continuum imagesA.5).

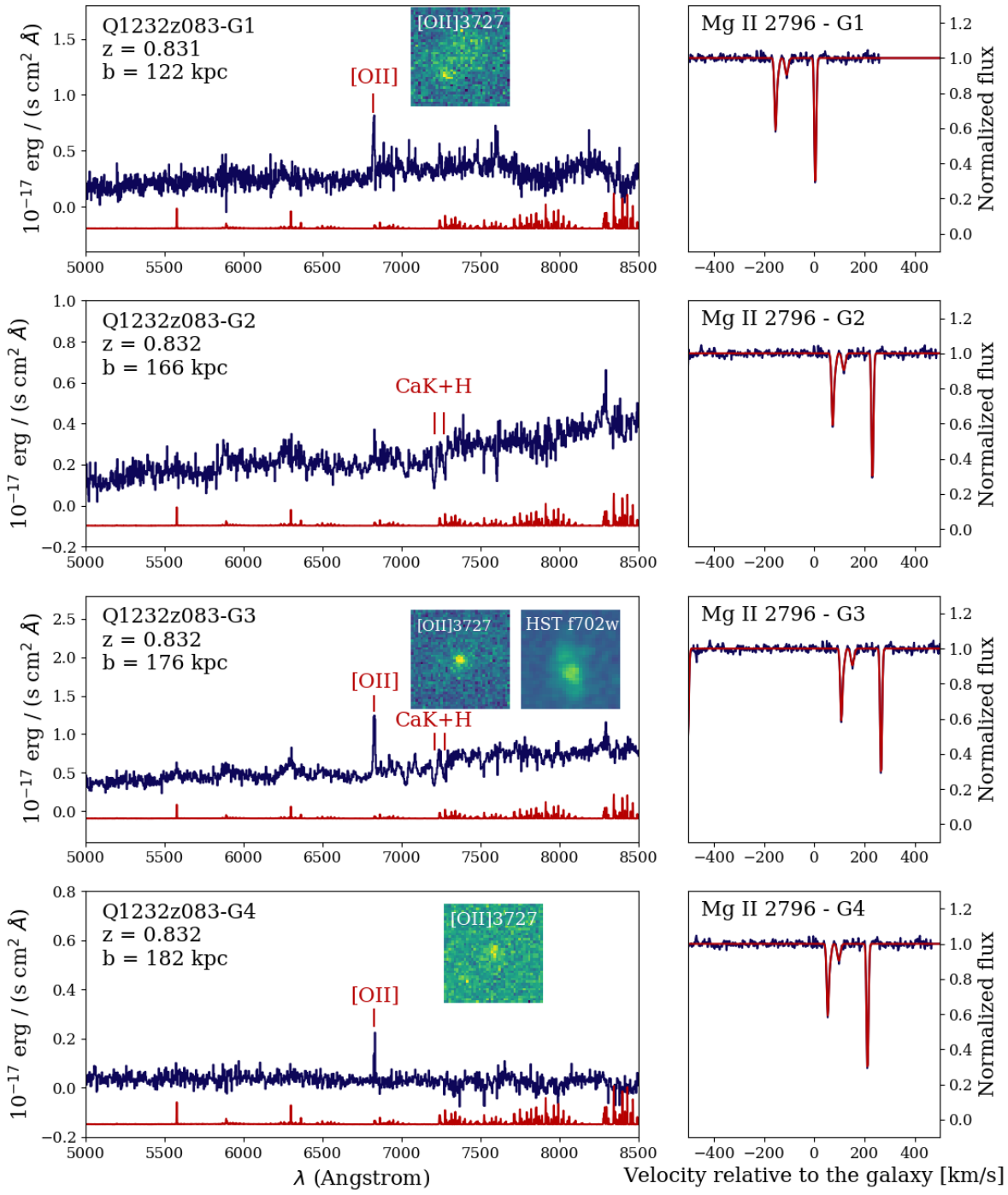

Figure A.15: Q1232z083 absorber and associated galaxies G1 and G2 were not detected in continuum in HST image. G1 is very diffuse object, G2 is and absorption line system, not detected in continuum in HST image, inset presents only MUSE white light image centered at the galaxy (continuum images A.5).

Bibliography

Aravena M., et al., 2019, ApJ, 882, 136

Arrigoni Battaia F., Hennawi J. F., Prochaska J. X., Oñorbe J., Farina E. P., Cantalupo S., Lusso E., 2019, MNRAS, 482, 3162

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Augustin R., et al., 2018, MNRAS, 478, 3120

Augustin R., et al., 2019, MNRAS

Bacon R., et al., 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 62690J (`arXiv:astro-ph/0606329`), doi:10.1117/12.669772

Bacon R., Piqueras L., Conseil S., Richard J., Shepherd M., 2016, MPDAF: MUSE Python Data Analysis Framework, Astrophysics Source Code Library (ascl:1611.003)

Bahcall N. A., 1996, arXiv e-prints, pp astro–ph/9611148

Barcons X., et al., 2012, arXiv e-prints, p. arXiv:1207.2745

Belfiore F., et al., 2017, MNRAS, 469, 151

Bergeron J., Boissé P., 1991, A&A, 243, 344

Bergeron J., Boulade O., Kunth D., Tytler D., Boksenberg A., Vigroux L., 1988, A&A, 191, 1

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bethermin M., et al., 2020, arXiv e-prints, p. arXiv:2002.00962

Bielby R., Crighton N. H. M., Fumagalli M., Morris S. L., Stott J. P., Tejos N., Cantalupo S., 2017, MNRAS, 468, 1373

Bielby R. M., et al., 2019, MNRAS, 486, 21

Boisse P., Le Brun V., Bergeron J., Deharveng J.-M., 1998, A&A, 333, 841

Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207

Bonato M., et al., 2018, MNRAS, 478, 1512

Boogaard L. A., et al., 2019, ApJ, 882, 140

Bordoloi R., et al., 2011, ApJ, 743, 10

Borisova E., et al., 2016, ApJ, 831, 39

Bouché N., Murphy M. T., Péroux C., Csabai I., Wild V., 2006, MNRAS, 371, 495

Bouché N., Murphy M. T., Péroux C., Davies R., Eisenhauer F., Förster Schreiber N. M., Tacconi L., 2007, ApJ, 669, L5

Bouché N., Hohensee W., Vargas R., Kacprzak G. G., Martin C. L., Cooke J., Churchill C. W., 2012, MNRAS, 426, 801

Bouché N., Carfantan H., Schroetter I., Michel-Dansac L., Contini T., 2015, GalPaK 3D: Galaxy parameters and kinematics extraction from 3D data (ascl:1501.014)

Bouché N., et al., 2016, ApJ, 820, 121

Cañas R., Elahi P. J., Welker C., del P Lagos C., Power C., Dubois Y., Pichon C., 2019, MNRAS, 482, 2039

Cantalupo S., Arrigoni-Battaia F., Prochaska J. X., Hennawi J. F., Madau P., 2014, Nature, 506, 63

Carilli C. L., Walter F., 2013, ARA&A, 51, 105

Chambers K. C., et al., 2016, arXiv e-prints, p. arXiv:1612.05560

Chen H.-W., Wild V., Tinker J. L., Gauthier J.-R., Helsby J. E., Shectman S. A., Thompson I. B., 2010, ApJ, 724, L176

Christensen L., Wisotzki L., Roth M. M., Sánchez S. F., Kelz A., Jahnke K., 2007, A&A, 468, 587

Christensen L., Møller P., Fynbo J. P. U., Zafar T., 2014, MNRAS, 445, 225

Christensen L., et al., 2017, A&A, 608, A84

Corlies L., Peeples M. S., Tumlinson J., O'Shea B. W., Lehner N., Howk J. C., O'Meara J. M., 2018, arXiv e-prints, p. arXiv:1811.05060

Daddi E., et al., 2010, ApJ, 713, 686

Daddi E., et al., 2015, A&A, 577, A46

De Cia A., 2018, A&A, 613, L2

Decarli R., et al., 2016, ApJ, 833, 69

Decarli R., et al., 2019, ApJ, 882, 138

Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, IEEE Proceedings, 97, 1482

Driver S. P., et al., 2019, The Messenger, 175, 46

Eisenhauer F., et al., 2003, in Iye M., Moorwood A. F. M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4841, Proc. SPIE. pp 1548–1561 (arXiv:astro-ph/0306191), doi:10.1117/12.459468

Ekers R. D., Massardi M., Sadler E. M., 2007, in From Planets to Dark Energy: the Modern Radio Universe. p. 47

Elahi P. J., Welker C., Power C., Lagos C. d. P., Robotham A. S. G., Cañas R., Poulton R., 2018, MNRAS, 475, 5338

Feldmann R., Gnedin N. Y., Kravtsov A. V., 2012, ApJ, 758, 127

Ferland G. J., et al., 2013, Rev. Mexicana Astron. Astrofis., 49, 137

Fossati M., et al., 2019, MNRAS, 490, 1451

Fox A., Davé R., 2017, Gas Accretion onto Galaxies. Vol. 430, doi:10.1007/978-3-319-52512-9,

Frank S., et al., 2012, MNRAS, 420, 1731

Fraternali F., 2017, in Fox A., Davé R., eds, Astrophysics and Space Science Library Vol. 430, Gas Accretion onto Galaxies. p. 323 (arXiv:1612.00477), doi:10.1007/978-3-319-52512-9˙14

Freundlich J., et al., 2019, A&A, 622, A105

Fumagalli M., O'Meara J. M., Prochaska J. X., 2016, MNRAS, 455, 4100

Fynbo J. P. U., et al., 2010, MNRAS, 408, 2128

Fynbo J. P. U., et al., 2013, MNRAS, 436, 361

Fynbo J. P. U., et al., 2018, MNRAS, 479, 2126

Gehrels N., 1986, ApJ, 303, 336

Genzel R., et al., 2010, MNRAS, 407, 2091

Genzel R., et al., 2012, ApJ, 746, 69

González-López J., et al., 2017, A&A, 608, A138

Hafen Z., et al., 2017, MNRAS, 469, 2292

Hamanowicz A., et al., 2020, MNRAS, 492, 2347

Hashimoto T., et al., 2018, Nature, 557, 392

Hemmati S., Yan L., Diaz-Santos T., Armus L., Capak P., Faisst A., Masters D., 2017, ApJ, 834, 36

Ho I. T., 2016, PhD thesis, University of Hawai'i at Manoa

Hodge J. A., da Cunha E., 2020, arXiv e-prints, p. arXiv:2004.00934

Hu W., et al., 2019, MNRAS, 489, 1619

Husemann B., Bennert V. N., Scharwächter J., Woo J. H., Choudhury O. S., 2016, MNRAS, 455, 1905

Husemann B., et al., 2019, ApJ, 879, 75

Jones T., Ellis R. S., Richard J., Jullo E., 2013, ApJ, 765, 48

Jones T., et al., 2015, AJ, 149, 107

Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2

Kacprzak G. G., Murphy M. T., Churchill C. W., 2010, MNRAS, 406, 445

Kacprzak G. G., Churchill C. W., Evans J. L., Murphy M. T., Steidel C. C., 2011, MNRAS, 416, 3118

Kacprzak G. G., Churchill C. W., Nielsen N. M., 2012, ApJ, 760, L7

Kanekar N., Neeleman M., Prochaska J. X., Ghosh T., 2018, MNRAS, 473, L54

Kennicutt Robert C. J., 1992, ApJ, 388, 310

Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531

Klitsch A., Péroux C., Zwaan M. A., Smail I., Oteo I., Biggs A. D., Popping G., Swinbank A. M., 2018, MNRAS, 475, 492

Klitsch A., et al., 2019a, MNRAS, 482, L65

Klitsch A., et al., 2019b, MNRAS, 490, 1220

Klitsch A., et al., 2020, MNRAS,

Kobulnicky H. A., Kennicutt Robert C. J., Pizagno J. L., 1999, ApJ, 514, 544

Krogager J. K., et al., 2015, ApJS, 217, 5

Krogager J. K., Møller P., Fynbo J. P. U., Noterdaeme P., 2017, MNRAS, 469, 2959

Krumholz M. R., Thompson T. A., 2007, ApJ, 669, 289

Kulkarni V. P., Cashman F. H., Lopez S., Ellison S. L., Som D., José Maureira M., 2019, arXiv e-prints, p. arXiv:1910.10759

Lagache G., Cousin M., Chatzikos M., 2018, A&A, 609, A130

Lagos C. d. P., Bayet E., Baugh C. M., Lacey C. G., Bell T. A., Fanidakis N., Geach J. E., 2012, MNRAS, 426, 2142

Lagos C. d. P., et al., 2015, MNRAS, 452, 3815

Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573

Lagos C. d. P., et al., 2019, MNRAS, 489, 4196

Lah P., et al., 2007, MNRAS, 376, 1357

Lane W. M., Briggs F. H., Turnshek D. A., Rao S. M., 1998, in American Astronomical Society Meeting Abstracts. p. 04.09

Lanzetta K. M., Bowen D., 1990, ApJ, 357, 321

Larkin J., et al., 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 62691A, doi:10.1117/12.672061

Le Brun V., Bergeron J., Boisse P., Deharveng J. M., 1997, A&A, 321, 733

Leethochawalit N., Jones T. A., Ellis R. S., Stark D. P., Richard J., Zitrin A., Auger M., 2016, ApJ, 820, 84

Lehner N., et al., 2013, ApJ, 770, 138

Lenkić L., et al., 2020, AJ, 159, 190

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782

Leroy A. K., et al., 2011, ApJ, 737, 12

Liu D., et al., 2019, ApJ, 887, 235

Lofthouse E. K., et al., 2020, MNRAS, 491, 2057

Lusso E., et al., 2019, MNRAS, 485, L62

Ly C., et al., 2007, ApJ, 657, 738

Mackenzie R., et al., 2019, MNRAS, 487, 5070

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Martin C. L., Shapley A. E., Coil A. L., Kornei K. A., Bundy K., Weiner B. J., Noeske K. G., Schiminovich D., 2012, ApJ, 760, 127

Matsuda Y., Nagao T., Iono D., Hatsukade B., Kohno K., Tamura Y., Yamaguchi Y., Shimizu I., 2015, MNRAS, 451, 1141

McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

Ménard B., Scranton R., Fukugita M., Richards G., 2010, MNRAS, 405, 1025

Møller P., Warren S. J., 1993, A&A, 270, 43

Møller P., Fynbo J. P. U., Fall S. M., 2004, A&A, 422, L33

Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691

Muratov A. L., et al., 2017, MNRAS, 468, 4170

Muzahid S., Kacprzak G. G., Charlton J. C., Churchill C. W., 2016, ApJ, 823, 66

Muzahid S., et al., 2020, MNRAS,

Narayanan D., Krumholz M. R., Ostriker E. C., Hernquist L., 2012, MNRAS, 421, 3127

Neeleman M., et al., 2016, ApJ, 820, L39

Neeleman M., Kanekar N., Prochaska J. X., Christensen L., Dessauges-Zavadsky M., Fynbo J. P. U., Møller P., Zwaan M. A., 2018, ApJ, 856, L12

Nelson D., et al., 2019, MNRAS, 490, 3234

Nielsen N. M., Churchill C. W., Kacprzak G. G., Murphy M. T., 2013, ApJ, 776, 114

Nielsen N. M., Kacprzak G. G., Pointon S. K., Churchill C. W., Murphy M. T., 2018, ApJ, 869, 153

Oteo I., Zwaan M. A., Ivison R. J., Smail I., Biggs A. D., 2016, ApJ, 822, 36

Oteo I., Zwaan M. A., Ivison R. J., Smail I., Biggs A. D., 2017, ApJ, 837, 182

Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., Katz N., Weinberg D. H., 2014, ApJ, 786, 54

Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2011, MNRAS, 410, 2237

Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2012, MNRAS, 419, 3060

Péroux C., et al., 2016, MNRAS, 457, 903

Péroux C., et al., 2017, MNRAS, 464, 2053

Péroux C., et al., 2019, MNRAS, 485, 1595

Planck Collaboration et al., 2016, A&A, 594, A13

Popping G., van Kampen E., Decarli R., Spaans M., Somerville R. S., Trager S. C., 2016, MNRAS, 461, 93

Popping G., et al., 2019, ApJ, 882, 137

Prochaska J. X., et al., 2017, ApJ, 837, 169

Queyrel J., et al., 2012, A&A, 539, A93

Quiret S., et al., 2016, MNRAS, 458, 4074

Rahmani H., et al., 2016, MNRAS, 463, 980

Rahmani H., et al., 2018a, MNRAS, 474, 254

Rahmani H., et al., 2018b, MNRAS, 480, 5046

Rahmati A., Schaye J., 2014, MNRAS, 438, 529

Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610

Rao S. M., Belfort-Mihalyi M., Turnshek D. A., Monier E. M., Nestor D. B., Quider A., 2011, MNRAS, 416, 1215

Rao S. M., Turnshek D. A., Sardane G. M., Monier E. M., 2017, MNRAS, 471, 3428

Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, MNRAS, 435, 2693

Rhee J., Lah P., Briggs F. H., Chengalur J. N., Colless M., Willner S. P., Ashby M. L. N., Le Fèvre O., 2018, MNRAS, 473, 1879

Riechers D. A., et al., 2019, ApJ, 872, 7

Rubin K. H. R., Prochaska J. X., Koo D. C., Phillips A. C., 2012, ApJ, 747, L26

Saintonge A., et al., 2017, ApJS, 233, 22

Sánchez S. F., et al., 2014, A&A, 563, A49

Schroetter I., et al., 2016, ApJ, 833, 39

Schroetter I., et al., 2019, MNRAS, 490, 4368

Schruba A., et al., 2011, AJ, 142, 37

Scoville N., et al., 2017, ApJ, 837, 150

Serra P., Jurek R., Flöer L., 2012, PASA, 29, 296

Serra P., et al., 2015, MNRAS, 448, 1922

Shull J. M., Danforth C. W., Tilton E. M., 2014, ApJ, 796, 49

Starck J. L., Pantin E., Murtagh F., 2002, PASP, 114, 1051

Steidel C. C., Dickinson M., Persson S. E., 1994, ApJ, 437, L75

Steidel C. C., et al., 2014, ApJ, 795, 165

Swinbank A. M., et al., 2012, MNRAS, 427, 1066

Tacconi L. J., et al., 2010, Nature, 463, 781

Tacconi L. J., et al., 2018, ApJ, 853, 179

Tacconi L. J., Genzel R., Sternberg A., 2020, arXiv e-prints, p. arXiv:2003.06245

Thatte N., et al., 2010, in Proc. SPIE. p. 77352I, doi:10.1117/12.857445

Theuns T., Srianand R., 2006, in Whitelock P., Dennefeld M., Leibundgut B., eds, IAU Symposium Vol. 232, The Scientific Requirements for Extremely Large Telescopes. pp 464–471 (`arXiv:astro-ph/0601637`), doi:10.1017/S1743921306001141

Tielens A. G. G. M., 2005, The Physics and Chemistry of the Interstellar Medium

Tumlinson J., et al., 2013, ApJ, 777, 59

Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389

Umehata H., et al., 2019, Science, 366, 97

Villanueva V., et al., 2017, MNRAS, 470, 3775

Walch S., et al., 2015, MNRAS, 454, 238

Walter F., et al., 2016, ApJ, 833, 67

Weilbacher P. M., Streicher O., Palsa R., 2016, MUSE-DRP: MUSE Data Reduction Pipeline (ascl:1610.004)

Whiting M. T., 2012, MNRAS, 421, 3242

Whiting M. T., Webster R. L., Francis P. J., 2006, MNRAS, 368, 341

Wisotzki L., et al., 2018, Nature, 562, 229

Wotta C. B., Lehner N., Howk J. C., O'Meara J. M., Oppenheimer B. D., Cooksey K. L., 2019, ApJ, 872, 81

Yanny B., Hamilton D., Schommer R. A., Williams T. B., York D. G., 1987, ApJ, 323, L19

Yanny B., York D. G., Gallagher J. S., 1989, ApJ, 338, 735

Yanny B., York D. G., Williams T. B., 1990a, ApJ, 351, 377

Yanny B., Barden S., Gallagher John S. I., York D. G., 1990b, ApJ, 352, 413

York D. G., et al., 2006, MNRAS, 367, 945

Zabl J., et al., 2019, MNRAS, 485, 1961

Zafar T., Popping A., Péroux C., 2013, A&A, 556, A140

Zuckerman B., Evans N. J. I., 1974, ApJ, 192, L149

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

de Jong R., 2019, in Preparing for 4MOST. p. 1, doi:10.5281/zenodo.3244904

Acknowledgements

Finally, my great dream came true - I became a Doctor in astrophysics. These past 3 years were a crazy ride and the two-fronts-battle: not only with data and PhD challenges but also with myself. Altogether I came from that experience stronger and wiser and I look forward to the new adventures ahead. This all would not be possible without a great number of people, who helped me, supervised and supported me along the way.

I would like to thank my PhD committee: prof. Volker Springel, prof. Simon White, Dr Klaus Dolag and Prof. Jörg Schreiber for their time and the nice atmosphere on the exam (big thanks to Prof. Schreiber for making the exam possible). Additionally, special thanks to Prof. Simon White for overseeing my PhD, all his valuable comments and making sure everything is proceeding in order.

Most of all biggest thanks to Dr Martin Zwaan and Dr Céline Péroux for your everyday supervision, for entrusting me with these interesting and demanding projects. Thank you that you believed in me despite all ups and downs. Thank you Anna and Dominika, you were my great support as fellow mentors throughout my time in ESO. Thank you for your guidance in the maze of the academia

ESO is a fantastic community where I felt very much at home and it is so sad to leave. While it is difficult to name all the fantastic people I met along the way, I would like to especially thank a few. Anne and Ramona my great Astro-sisters always here to support, help and guide: Anne - how would I get not know ALMA without your help? Ramona - the world of QSO absorbers, HST and doggos you showed me got me a new job and lots of joy. Katja - thank you for all talks, food, drinks, lake and keeping examining my knowledge before the defence. Adrian, thank you for keeping up with my annoying self during countless climbing sessions and hiking trips. And I cannot forget about the angel Nelma! There is no issue with cannot be resolved bu that genius, thank you!

Great thanks and hugs for all parties, vacations, food, laughs, trips, hikes swims etc. comes also to Melanie, Sara, Shweta, Misa, Monika, Roland, Johanna, Maria Gullia, Carlo, Calum, Tereza, Nelly & Pablo, Josh and countless others as well as my dear Annas with who could talk in Polish and explore both Munich and Heidelberg. To my band: Georgia, Thomas Maria (and Viola!) for making me believe in rock-n-roll again and helping me stay sane with all the music.

Szczególne podziekowania dla mojej rodziny: moich ukochanych rodziców którzy wspierali mnie na każdym kroku i nie pozwolili sie poddać a w każdej potrzebie byli w gotowości. Moim kochanym dziadkom, wiem że jestescie dumni z pani doktor w rodzinie! Kocham

was bardzo i dziekuje za pomoc i wyrozumiałość.

And last but not least to all of you fantastic, kind people who make ESO such a wonderful and friendly place, thank you for every day, every chat and every coffee. Science makes our life paths very complicated but the intersections are plenty so I am sure we will meet each other again!